Measuring value-added in noncognitive learning outcomes in higher education institutions: A civic engagement perspective

Author: Yang Wang

BOSTON COLLEGE

Lynch School of Education

Department of
Educational Research, Measurement, and Evaluation

MEASURING VALUE-ADDED IN NONCOGNITIVE LEARNING OUTCOMES IN HIGHER EDUCATION INSTITUTIONS: A CIVIC ENGAGEMENT PERSPECTIVE

Dissertation
by

YANG (CAROLINE) WANG

submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

May 2012

MEASURING VALUE-ADDED IN NONCOGNITIVE LEARNING OUTCOMES IN HIGHER EDUCATION INSTITUTIONS: A CIVIC ENGAGEMENT PERSPECTIVE

by

Yang (Caroline) Wang

Dr. Henry I. Braun, Dissertation Chair

ABSTRACT

Addressing the call to provide hard evidence on undergraduate student outcomes and make comparisons across higher education institutions for accountability purposes, this study extends current efforts in measuring higher education outcomes and explores the differences in three value-added methodologies. Using the CIRP freshman and senior survey data from 2002 and 2006, this study examines noncognitive higher education outcomes with a focus on civic engagement. The three value-added methodologies examined are: an OLS-based cross-sectional method, an HLM-based cross-sectional method, and an HLM-based longitudinal method. Rather than seek to establish which methodology is superior, this study intends to provide empirical evidence concerning the similarities and differences in estimating institutional effectiveness with regard to civic engagement. First, several student-level and institution-level covariates were found to be associated with a measure of civic engagement in the senior year after adjusting for their level of civic engagement as freshmen. The model comparison further revealed some advantages in the HLM-based longitudinal method over the other two methods, such as providing a more accurate institutional value-added estimate and the ability to account for

a relatively large percent of the total variance in the civic engagement measure when using the same covariates. Next, among all pairs of model comparisons, results from the two HLM-based methods agreed the most (r = .80). However, institutional rankings fluctuate dramatically, even when comparing institutions within small peer groups. Finally, the findings highlighted great divergences among different value-added methodologies in identifying institutions that perform significantly differently from the average for accountability purposes.

ACKNOWLEDGMENTS

There are many people I have met through my life, without whom I would never have reached this point or completed this dissertation. I am thankful for all of them and cannot possibly name everybody. Among them and their more notable contributions are:

My dissertation chair Dr. Henry Braun and other committee members, Dr. Laura O'Dwyer, Dr. Karen Arnold, are especially important. Their feedback, advice, and overall support buoyed me during this time. Dr. Braun's input and careful reading of the many drafts kept me focused on the details of the project. Dr. O'Dwyer's support through this process and during most of my time at Boston College has kept me on the right side of the border between sane and insane. Dr. Arnold's insight from the higher education field was essential for me to understand many of the issues addressed here.

The rest of the faculty and staff in the Lynch School's ERME program and Boston College has made my time here a positive one. From Dr. Michael Russell's enthusiastic welcome when I first walked onto BC's campus to Dr. Larry Ludlow's continued encouragement to Dr. Mandy Li's inspiration in the later stages of my studies, all have contributed to my development. Moreover, I wish to thank Fr. James Fleming, who inspired me to pursue survey research in the field of higher education.

Many people were involved in emboldening and supporting me in my prior life. My mentor at BLCU, Dr. Xiaoqing Xie, provided me with the foundational knowledge I needed. My colleagues and supervisors at the Personnel Testing Center in Beijing provided me the confidence that I could succeed here.

Finally, my family and friends are those closest to me though my life and have put up with me at my worst. Yet, I have been able to rely on them without fail. A special thank you would go to my husband, for the love and encouragement during the years – I would not have accomplished this without you.

To all of the above and those still unnamed, I thank you for all that you have done to help me in this journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1. INTRODUCTION

1.1 Description of the Problem

During the last decade, holding schools and teachers accountable for student learning outcomes has gained unprecedented popularity in K-12 education (see No Child Left Behind Act, 2002; U.S. Department of Education, 2010; Wainer, 2004). With the rising costs of college and a sluggish economy, there is growing interest in providing similar hard evidence of student learning outcomes in higher education. Parents want to know whether their children will obtain the knowledge they need for their future and whether the substantial expenditures are worthwhile. Even President Obama has focused on this in his recently unveiled Blueprint for College Affordability, which includes an emphasis on higher education outcomes and reports to students and parents through a College Scorecard (Compton, 2012; Obama, 2012). Employers are also interested in the abilities of new college graduates entering the workforce. At the same time, administrators and program directors at postsecondary institutions need evidence to demonstrate how they are doing and what can be done to improve their institutions. Therefore, both external and internal stakeholders in the higher education community need to understand the quality of postsecondary education.

This urgency is further highlighted in the report from the Commission on the Future of Higher Education appointed by former U.S. Secretary of Education Margaret Spellings, in which accountability was stressed as one of the four central issues in higher education in addition to access, affordability, and quality (U.S. Department of Education, 2006). The report further indicated that higher education institutions "must become more

transparent" and "must willingly share" student success outcomes; "student achievement… must be measured by institutions on a 'value-added' basis that takes into account students' academic baseline when assessing their results", so that the information can be made available to "measure the relative effectiveness of different colleges and universities" (p.4). The Spellings Commission believed that accountability is vital to ensuring the success of the other three central issues examined, and urged access to hard evidence regarding student learning outcomes and the comparability of such evidence across higher education institutions.

The U.S. higher education community is not the only group concerned with how students learn and progress during college. The Organization for Economic Cooperation and Development (OECD) has launched the Assessment of Higher Education Learning Outcomes (AHELO) initiative to test student and university performance globally. Designed as "a test of university students comparable internationally," this assessment aims at understanding "what students in higher education **know and can do upon** graduation" in an international context (OECD, 2010-2011; OECD, 2011). It is hoped that AHELO will launch in 2016, with the timetable being dependent on the results of a feasibility study (OECD, 2011). A high quality postsecondary education system is a cornerstone to the development of the workforce and society. Therefore, evaluations of U.S. higher education should be cognizant of the larger global arena.

However, before evidence can be provided regarding undergraduate students' learning outcomes and the quality of higher education institutions, essential learning outcomes need to be conceptualized. Over the past 50 years, numerous theories and

models have been proposed for the dimensions of student learning outcomes in postsecondary education. Among all the educational outcomes identified, students' cognitive attributes, such as critical thinking, analytic reasoning, and writing measured by standardized assessments, have gained extensive attention in research and practice. On the other hand, the outcomes that are associated with noncognitive attributes, such as students' attitudes, values, and beliefs, are far less developed, partly because of its complexity in conceptualization and difficulty in measurement.

Although certainly important, research has shown that cognitive abilities are probably emphasized more than they should be (Cunha & Heckman, 2010; Levin, 2011). In meta-analysis and other research, cognitive attributes are found to explain only a small portion in the relationship between schooling attainments and economic outcomes, and a small portion of variance in adult success as measured by earnings and supervisory ratings in the workplace (Bowles, Gintis, & Osborne, 2001; Levin, 2011). On the other hand, noncognitive attributes are "important for explaining valuable individual and social outcomes including economic productivity" (Levin, 2011, p. 5). For example, after controlling for cognitive variables, noncognitive variables are found to offer an incremental prediction of earnings; however, the cognitive variables do not offer similar contributions if done in the reverse order (Cunha & Heckman, 2010). Therefore, as Levin (2011) suggested, cognitive abilities are not the dominant predictors when explaining social and economic outcomes. Consequently, the educational system should not solely rely on cognitive outcome measurements when measuring education quality.

Noncognitive attributes, as an indispensable component of higher education, comprise important elements and should be considered in institutional effectiveness studies.

Nevertheless, it is hard to identify key learning outcomes with unanimous agreement because each of the proposed theories and models regarding student learning outcome dimensions was developed from a unique perspective. However, with the start of this millennium, the Association of American Colleges and Universities (AAC&U) became active in leading national collaborations on a series of research projects exploring higher education outcomes (AAC&U, 2002, 2004, 2005, 2007, 2010). AAC&U's multiyear efforts have resulted in "**a remarkable consensus on a few key outcomes that all students, regardless of major or academic background, should achieve during undergraduate study**" (AAC&U, 2005, p.2). Beyond general knowledge and practical skills, one key outcome associated with students' noncognitive attributes is civic knowledge and engagement at both the local and global levels. The importance of this as a focus in higher education is best summed up by AAC&U (2012) stating that "civic knowledge and capability are not bestowed at birth. They are hard won, through education at all levels" (p. 69).

Civic engagement is a broad term that refers to both inward and outward aspects of people's propensity of being an active member in a civil society (see AAC&U, 2010; Ehrlich, 2000). Here, the inward aspect can be demonstrated from one's commitment and adjustment in attitudes and beliefs towards civic involvement. The outward aspect includes various forms of participation in civic activities, such as voting, volunteering, and involvement in community service and social movements. The engagement of its

citizenry in the community is a fundamental component of a democratic society (Dewey, 1916; Putnam, 2000). Preparing young adults to become responsible citizens not only strengthens a democratic society, it also benefits both the public and individuals. Higher education institutions, as the final stage of formal education for many Americans, have a longstanding goal of fostering civic responsibilities among college students (Bowen, 1977; Colby, Ehrlich, Beaumont, & Stephens, 2003; Hurtado, 2007; Morse, 1989). Even in the earliest days of higher education in the U.S., fostering civically engaged students was a concern. This can be seen in the 1701 Yale University Charter, whose purpose is stated as developing students who "may be fitted for Publick employment both in Church and Civil State" (Yale University, 1976). Today, American colleges and universities, regardless of their diverse education missions, all have a common goal that speaks to the cultivation of active and civically engaged citizens (Ehrlich, 2000; Pew Partnership for Civic Change, 2004).

Since U.S. colleges and universities have been educating students along multiple dimensions involving academic disciplines as well as moral/ethical instruction (Reuben, 1996), collected higher education outcomes should include not only students' cognitive academic performance, but also their noncognitive development on constructs such as civic engagement. As Newman (1985) emphasized in a Carnegie Foundation's special report:

> If there is a crisis in education in the United States today, it is less that test scores have declined than it is that we have failed to provide the education

for citizenship that is still that most important responsibility of the nation's schools and colleges. (p.31)

However, although educating civically engaged citizens is a critical goal of higher education institutions, measurement of students' development in civic engagement is largely overlooked at the postsecondary level.

Civic engagement is the focus of this study because it is regarded as one of the most important outcomes of education by many people. As Torney-Purta and Amadeo (2011) stated, studies of civic engagement contributes at the macro level, the meso level, and the micro level.[1] Those studies provide a multidimensional view of education and how well young people grow in general. They satisfy the need to communicate such information to the public. They also provide critical information to teachers in the classrooms and to students themselves. However, this study investigates the field of civic engagement, not only because of civic engagement's critical role in democratic society, but also because of the challenges to measuring civic engagement in higher education.

To measure college students' learning progress and evaluate the effectiveness of higher education institutions requires large longitudinal databases that track students from multiple institutions over the course of their college years. Information collected in these datasets should include not only students' academic performance, but also their noncognitive development and various factors associated with student learning, such as background characteristics, high school performance, college experiences, and

[1] Torney-Purta and Amadeo (2011) define macro level as the "multidimensional picture of countries' education systems and to the diverse cultural settings in which young people develop"; meso level as the "education policy specialists" and "those who communicate information about education to the public such as journalists," and "specialists in test development"; micro level as "teachers and students in the classroom" (p.2).

institutional characteristics. Unlike in the K-12 setting where large-scale longitudinal datasets with several of these critical features have been collected and used in daily practice and research for years, such longitudinal datasets have not been available in the field of higher education. In reality, there has been limited national focus on the evidence of student learning outcomes in higher education. Furthermore, policy makers, educators, and researchers all have difficulties with the little attention given to measuring postsecondary student learning outcomes (e.g., Anderson, 2011; Banta, 2008; Jones, 2009).

Since measuring value-added in higher education institutions is another focus of this dissertation, a recent collaborative effort using value-added techniques to form a nationwide data and accountability system in the U.S. higher education is briefly introduced here. In 2007, two leading organizations in U.S. higher education – the American Association of State Colleges and Universities (AASCU) and the National Association of State Universities and Land-Grant Colleges (NASULGC) – initiated a Voluntary System of Accountability℠ (VSA) program to measure core educational outcomes in higher education and to provide "clear, accessible, and comparable information on the undergraduate student experience to important constituencies through a common web report – the College Portrait" (VSA, n.d.). Three standardized outcome instruments were selected to measure these core education outcomes – the Collegiate Assessment of Academic Proficiency (CAAP, from ACT®), the ETS® Proficiency

Profile[2] (from the Educational Testing Service), and the Collegiate Learning Assessment (CLA, from the Council for Aid to Education, CAE). The VSA program has attracted nationwide attention since its inception. As of July 2011, 319 institutions from 48 states have participated in the VSA initiative (VSA, 2011). As the VSA program and its College Portrait provide transparent and comparable information about all participating institutions to various stakeholders and the general public, it is the best effort to date demonstrating student learning outcomes in college.

However, as a work in progress, there are many areas where VSA can improve. First, as education at the postsecondary stage involves multidimensional values, multiple instruments, rather than a single assessment, are needed to measure student college outcomes. Currently, VSA's measure of institutional effectiveness is only based on one of the three selected standardized assessments which measure student cognitive learning in critical thinking, analytic reasoning, and written communication. Noncognitive outcomes such as individual responsibility, social responsibility, and teamwork and leadership skills that are "highly prized both by the academy and by employers," are largely ignored (AAC&U, 2005, p.1). Although VSA suggests institutions provide "campus learning climate data" through student experience surveys, results from these noncognitive outcome surveys have not been used in the value-added measure (McPherson & Shulenburger, 2006). While there is a legitimate reason and a great need for the information, various stakeholders in the higher education system still lack this necessary information for making decisions and judgments about particular institutions.

[2] The assessment was renamed in 2009. Its former name was the Measure of Academic Proficiency and Progress (MAPP).

Another area where VSA may be able to improve is its study design. Although the advantages of a true longitudinal study over a repeated cross-sectional study are well-known, VSA adopts a "quicker, simpler, and less costly" cross-sectional study design to measure student progress and estimate institutional effectiveness due to practical considerations (VSA, 2008a, p.4). Specifically, it measures freshman and senior students attending the same institution in the same year, rather than following the same group of students over time. This cross-sectional design forms the basis of many critiques of VSA's measure of institutional value-added (see Larson, 2008; Garcia, 2007; Pike, 2006). Originally, all three test agencies used an ordinary least squares (OLS) based method to estimate institutional value-added scores (ACT, 2009; Liu, 2008; VSA, 2008a). Starting from the 2009-2010 assessment cycle, CLA introduced an advanced estimation method using a hierarchical linear modeling technique (Steedle, 2009). Therefore, different value-added methods are now used to analyze institutional value-added depending on the specific test each institution chooses to use. Many concerns exist, but there is little empirical evidence addressing the differences in analyzing higher education learning outcomes among these two methods based on cross-sectional study designs and other methods based on longitudinal designs.

Higher education has long been found to have a profound impact not only on individual students' academic and cognitive development, but also in students' noncognitive development (Astin, 1993a; Pascarella & Terenzini, 1991, 2005). Student learning outcomes have multiple dimensions. Cognitive development, such as critical thinking, analytic reasoning, and writing measured by standardized assessments,

represents only one aspect of students' advancements during college years. Noncognitive attributes, as an indispensable component of higher education, have potential and should be considered.

Although VSA has put great effort into measuring broad cognitive skills, there is not much knowledge of, and research into, measuring noncognitive learning outcomes in higher education. In addition, it is not clear whether the two cross-sectional estimation methods currently used by VSA and the longitudinal estimation method widely accepted in K-12 would find different results in estimating institutional effectiveness. Nor is it clear whether the value-added methodology used to analyze cognitive standardized test scores can be used with noncognitive survey measures.

1.2 Purpose and Research Questions

To help bring clarity to these issues, this study explores the problems and uncertainties in measuring institutional value-added in higher education by focusing on value-added methodology issues and on student noncognitive learning outcomes. Here, the purpose of value-added modeling is to develop indicators of the effects of higher education institutions on students' learning outcomes while taking into account students' prior achievement and other associated factors (e.g., background characteristics) – both at the student and the institutional level.

In this study, the estimated institutional value-added scores refer to the differences (or residuals) between the observed and the expected institutional mean outcomes scores. For example, when data are available from the same student at two time points – the

freshman year and the senior year, institutional value-added scores are the average differences between observed and expected senior students' outcome scores after adjusting for those students' scores in their freshman year, as well as other student characteristics. Strictly speaking, value-added models in this study do not model score gains (i.e., senior scores - freshman scores) or growth rate (i.e., senior scores/freshman scores). The substantive interest is to model senior students' status scores with an adjustment of scores from the freshman year. Therefore, although the term *growth* is loosely used as a convention in many studies (e.g., CAE, 2010; Hershberg, Simon, & Lea-Kruger, 2004; LaNasa, Olson, & Alleman, 2007), this study avoids the use of the term. This concept is revisited in later sections (e.g., section 3.3.4).

It should be noted that the institutional value-added estimations are not developed to rank institutions. The estimated value-added scores are more useful in comparing schools with similar student bodies and those who have similar programs. For example, it is reasonable to compare a group of Catholic schools which are highly selective and focus on undergraduate liberal arts education. Yet, it may not be as valuable to compare a large public university which has comprehensive curriculum offerings and a diverse student body with a small urban community college which has limited programs of study and a predominantly urban student body.

Using survey data from the Cooperative Institutional Research Program (CIRP) at the Higher Education Research Institute (HERI) in the University of California, Los Angeles and the Integrated Postsecondary Education Data System (IPEDS) from the National Center for Education Statistics (NCES), students' development in the

noncognitive outcome of civic engagement is explored. The research questions of the study are threefold. The first and second research questions aims to identify student and institutional characteristics associated with the noncognitive outcome. The findings help ensure the credibility of estimating institutional value-added in the third research question where longitudinal and cross-sectional methodologies are compared. Moreover, issues associated with using student learning outcomes as an indicator of institutional effectiveness and the implications of using assessment-as-accountability measures for educators and policy makers are discussed. The three guiding research questions are as follows:

1. To what extent are undergraduate students' characteristics associated with a measure of civic engagement in the senior year after adjusting for their level of civic engagement as freshmen?

2. To what extent are institutional characteristics (such as school type, selectivity, location, and type of location) associated with a measure of senior students' civic engagement after adjusting for their level of civic engagement as freshmen, as well as their characteristics?

3. What are the similarities and differences among the results of longitudinal and cross-sectional value-added models in calculating higher education institutions' value-added scores with regard to civic engagement?

1.3 Significance of the Study

There has been an urgent call for a culture of evidence among higher education communities. This call for evidence comes from various external and internal stakeholders for consumer information and self-reflection. The Spellings commission conveyed this call in *A Test of Leadership* (U.S. Department of Education, 2006):

> Among the vast and varied institutions that make up U.S. higher education, we have found much to applaud but also much that requires urgent reform. (p.vi) … To meet the challenges of the 21st century, higher education must change from a system primarily based on reputation to one based on performance. We urge the creation of a robust culture of accountability and transparency throughout higher education. (p.20)

Nationwide, this call for evidence has gained support from reputed scholars and empirical studies with books titled *Our Underachieving Colleges* (Bok, 2006) and *Academically Adrift* (Arum & Roksa, 2011). Worldwide, this call for evidence received an influential echo from OECD's AHELO – an international comparative initiative. Therefore, the first important result of this study would be to provide one perspective of the evidence needed. Policy makers and various educators will find the empirical findings helpful to understand what students have learned and whether they have progressed in civic engagement as it applies to both the overall patterns and different subgroups of institutions and students.

Second, most research focuses on using well-established cognitive assessments to measure higher education learning outcomes. Limited research has been done on

noncognitive outcomes. As most U.S. colleges and universities have a civic mission and take responsibility for preparing young adults to be responsible citizens in the future, this study focuses on measuring undergraduate students' development in civic engagement and responsibility – a noncognitive outcome that is critical for both the students themselves and democratic society. Since no study has investigated higher education institutions' value-added on civic engagement, this study is the first systematic study that delves into this issue and reveals the results to the stakeholders. Thus, it provides a unique contribution to all those who are interested in understanding students' noncognitive development – especially development in civic engagement.

Third, the field of measuring higher education outcomes is relatively new. Not only do uncertainties exist as to which outcomes should be measured, but also which methodology should be adopted to analyze the results. Using large-scale empirical data, this study explores differences between cross-sectional and longitudinal methodologies in measuring higher education outcomes. Specifically, two estimation methods using cross-sectional designs and one estimation method using a longitudinal design are compared. Findings of this study will contribute to the developing literature on postsecondary institutional value-added methodology issues. Implications beyond the empirical evidence are presented. Additionally, researchers and practitioners in the field may find the study results helpful in selecting the appropriate value-added method for their own study.

Lastly, surveys normally rely on self-selected respondents and are limited in their ability to generalize the study results to other samples and situations. This study adopts

rigorous statistical analysis procedures, such as the use of confirmatory factor analysis and item response theory in constructing the civic engagement outcome scale and the use of post-stratification on survey respondents. These refined analyses help with producing more credible results than most other studies using survey data from self-selected respondents.

1.4 Outline of the Dissertation

After an introduction to the purpose and research questions of the dissertation study in this first chapter, chapter 2 reviews literature in the field of higher education related to three issues: (1) the multiple dimensions of college student learning outcomes; (2) the noncognitive dimension and the concept of civic engagement; and (3) current postsecondary outcome measures and challenges.

Chapter 3 introduces the research design of the study. It describes the data used in this study and presents the methods used to construct the civic engagement variable. Due to the importance of the outcome variable in later analyses, a special section is devoted to the explication of the three stages involved in constructing an indicator of this latent variable. In the last section, the statistical models for each research question are presented in detail.

In chapter 4, results from empirical data analyses are presented. The preliminary results include those from the stages of variable construction and descriptive analyses on the civic outcome variable, as well as other variables of interest. Results from each model are then presented to answer the research questions of the study.

Finally, chapter 5 provides further discussion of the research questions and implications of the issues related to value-added models and the effectiveness and accountability of higher education institutions. Policy-related recommendations are provided. Finally, the dissertation ends with a discussion of limitations of the study and future research directions in the field.

CHAPTER 2. LITERATURE REVIEW

This chapter provides an overview of research on higher education learning outcomes and indicates how this study relates to the broader discussions in this field. In doing so, this chapter is organized into three major sections. First, researchers' understandings of the dimensions of student learning outcomes in higher education are presented. Various student learning outcome taxonomies, student development theories, and recent collaborative work in this field are briefly introduced. A framework for the current study is then developed with the aim of avoiding potential conceptual confusion. The second section highlights the importance of noncognitive development in students' college experiences and, specifically, past research findings on the effect of college on students' formation of civic engagement. The third section focuses on what and how higher education learning outcomes are currently measured with an introduction of the three value-added estimation methods. It finishes with an overview of the challenges in measuring institutional effectiveness in higher education.

2.1 Dimensions of Student Learning Outcomes

What is meant by "student learning outcomes" in higher education? Before considering how to best measure it, the content and scope of student learning outcomes should be conceptualized. One characteristic of the American higher education system is its diversity. Because each institution has a unique mission, the simple task of conceptualization is not straightforward. This complexity is illustrated in the following two aspects.

One is the plethora of existing theories, models, and research related to college student learning, student development, and college outcomes. From seminal works and early student developmental theories (e.g., Astin, 1973; Chickering & Reisser, 1993; King & Kitchener, 1994; Perry, 1970) to more recent collaborative research, reports, and discussions (e.g., Adelman, Ewell, Gaston, & Schneider, 2011; AAC&U, 2004; NASPA & ACPA, 2004; Center of Inquiry, 2009; Dwyer, Millett, & Payne, 2006), various classifications of the outcomes of higher education have been proposed.

The other aspect of this complexity comes from the inconsistency of concepts in different fields or sometimes within the same field of the research literature. That is, the same construct can be expressed differently or different constructs may adopt similar names, creating conceptual confusion. The next section highlights different views of student learning outcome classifications. The summary then concludes with a framework to provide conceptual clarity.

2.1.1 Early Views of Outcome Taxonomies and Student Development

In the 1970's and 1980's, a variety of outcome taxonomies of higher educational student learning were developed for institutional planning and research. Among them, Astin (1973), Bowen (1977), and Lenning (1977) all provided comprehensive views of college outcomes from a relatively broad perspective. Meanwhile, Mentkowski and Doherty's (1984) taxonomy was developed by a group of scholars at Alverno College; therefore, it was more institutionally focused (Jacobi, Astin, & Ayala, 1987).

After an extensive review of literature on student outcomes, Bowen (1977) developed a five-category taxonomy from a practical point of view. These categories are closely connected with the goals many institutions hold – cognitive learning, emotional and moral development, practical competence, direct satisfactions from college, and the avoidance of negative outcomes. Although Lenning's (1977) taxonomy is also well-defined and comprehensive with five main categories, it was developed for the National Center for Higher Education Management Systems and, therefore, reflected a management perspective. The main categories include: (1) economic outcomes, (2) human characteristics outcomes, (3) knowledge, technology, and art form functions, (4) resource and service provision outcomes, and (5) aesthetic and cultural activities. Strictly speaking, student outcomes only comprise a portion of Lenning's taxonomy.

Astin's (1973) taxonomy, meanwhile, has had a strong influence on higher education literature. In his more complex taxonomy, postsecondary outcomes are organized along three dimensions: the type of outcome, the type of data, and the time span involved. The first two dimensions can be thought of as a 2 x 2 matrix where type of outcome consists of two broad domains in behavioral science – cognitive and noncognitive (or affective) and type of data consists of psychological and behavioral. The third dimension, time span, can be thought of as a continuous variable, since it could be four years of college or one, five, or ten years after graduation from college. The 2 x 2 taxonomy from the first two dimensions forms four categories of postsecondary outcomes: cognitive-psychological, cognitive-behavioral, affective-psychological, and affective-behavioral. Using this taxonomy as a guideline in organizing the evidence of

postsecondary outcomes, Pascarella and Terenzini (1991, 2005) wrote a comprehensive syntheses of *How College Affects Students*. Their books organize postsecondary outcomes into different chapters. Some chapters reflect one of the four cells in Astin's taxonomy. For example, two chapters in their books, *psychosocial change* (chapter five) and *attitudes and values* (chapter six) fall into the affective-psychological cell; the *educational attainment and persistence* chapter (chapter eight) falls into the cognitive-behavioral cell. Other chapters generally cover multiple cells. For example, the chapter on *moral development* (chapter seven) covers both the cognitive-psychological and affective-psychological cells.

Compared with the above mentioned outcome taxonomies, Mentkowski and Doherty's Alverno taxonomy has the most categories. Eight outcomes were identified which reflected faculty members' and administrators' views about the goals of the liberal arts institution – communications, analysis, problem solving, valuing, social interaction, taking responsibility for the environment, involvement in the contemporary world, and aesthetic response. This student outcome taxonomy provides an alternative perspective from one specific institution. Although developed in the mid-1980s, many current postsecondary institutions claim similar or overlapping education goals and concepts.

These four outcome taxonomies represent early views of higher education outcomes. They were developed from the need for institutional planning and management or with the aim of enhancing educational programs and achieving their ultimate missions and have been widely used in institutional planning and research. For example, the Alverno taxonomy has been used in the implementation of an outcome-

oriented liberal arts program at Alverno College and supported longitudinal studies of over two decades examining student learning there (Mentkowski & Associates, 2000). Other important taxonomies examine the issue from the perspective of student development – known as student development theories. Student development theories focus more on a systematic view of student characteristics and their change, growth, and development.

Emerging with the development of psychology, student development theory has a long history and has developed many well-known psychological and sociological theories and models. Two general types of student development theories exist – psychosocial theories and cognitive-structural theories (NASPA & ACPA, 2004).

Psychosocial theories focus on intrapersonal and interpersonal lives, and development occurs when students "…qualitatively change their thinking, feeling, behaving, valuing, and relating to others and oneself" (Chickering & Reisser, 1993 in Evans, Forney, & Guido-DiBrito, 1998, p. 10). One of the well-known psychosocial theories in this category is Chickering's seven vectors of student development (1969, 1993). Chickering's theory was the first major theory focusing on college students (Evans, Forney, Guido, Patton, & Renn, 2010). In the second edition of the landmark book *Education and Identity* (Chickering & Reisser, 1993), the definitions and ordering of the seven nonlinear and interrelated vectors are revised as: developing competence, managing emotions, moving through autonomy toward interdependence, developing mature interpersonal relationships, establishing identity, developing purpose, and developing integrity. These vectors emphasize identity formation through a student's

college years when multiple vectors usually evolve concurrently. In general, Chickering's theory provided a comprehensive picture of student development in higher education. It has generated extensive research and has had a major impact on understanding and describing college students' psychosocial development.

Cognitive-structural theories (also known as cognitive and moral development theories, cognitive theories) examine how students develop intellectually, and how they interpret the world around them. Kohlberg's (1969) theory of moral development and Perry's (1970, 1981) theory of intellectual and ethical development are two examples in this category. These theories examine *how* people think, rather than *what* people think (Evans et al., 2010). In other words, they represent continuous and cumulative developmental stages to explain how students interpret the world around them. For example, Perry's theory progresses from dualism (answers are either right or wrong, black or white), to multiplistic (multiple expert opinions exist), to relativistic (answers are right sometimes but not always), to commitment in relativism (nothing is always right for every situation; Perry, 1970, 1981). Partly because these concepts are easy to identify and recognize in college students, Perry's theory has been used extensively and is of great value in higher education settings. Kegan's (1994) theory of the evolution of consciousness and Baxter Magolda's (2001) theory of self-authorship are two theories developed more recently. Both focus on how people develop meaning and build constructs. They have also had important impact on classroom instruction and student affairs divisions in higher education (Evans et al., 2010).

Although developed decades ago, student outcome taxonomies and student development theories are the foundations of how student learning outcomes are conceptualized and still influence how this issue is addressed today. Since an extended explanation of these early taxonomies and theories is beyond the scope of this dissertation, several influential ones are briefly introduced here to provide a broader picture of the complexity of student outcome dimensions. The next section examines more recent studies and research.

2.1.2 Recent Collaborative Research

Recognizing the importance of high-quality higher education in today's global economy, an unprecedented collaboration among various stakeholders in the higher education community has developed. Many reports and research projects are being conducted among higher educational associations, national research centers, accreditation agencies, institutions, and workforce employers. Understanding the dimensions of student outcomes are the focus of many of these projects.

The Association of American Colleges and Universities is one of the leaders in this national collaboration. Leading a national initiative – Liberal Education and America's Promise (LEAP), AAC&U has facilitated a multiyear dialogue with hundreds of colleges and universities and has worked closely with business employers. A series of studies on liberal education outcomes (AAC&U, 2002, 2004, 2005, 2007) have been conducted to provide a new framework and guide for students going through the transitions of college. Specifically, AAC&U has organized a set of essential learning

outcomes that "**all students, regardless of major or academic background, should achieve during undergraduate study**" (AAC&U, 2005, p.2). These outcomes are grouped into four categories, which AAC&U summarized as follows:

(1) Knowledge of human cultures and the physical and natural world

- Through study in the sciences and mathematics, social sciences, humanities, histories, languages, and the arts

(2) Intellectual and practical skills, including

- Inquiry and analysis
- Critical and creative thinking
- Written and oral communication
- Quantitative literacy
- Information literacy
- Teamwork and problem solving

(3) Personal and social responsibility, including

- Civic knowledge and engagement – local and global
- Intercultural knowledge and competence
- Ethical reasoning and action
- Foundations and skills for lifelong learning

(4) Integrative learning, including

- Synthesis and advanced accomplishment across general and specialized studies

(AAC&U, 2007, p.3)

Drawn from collaborations of various constituencies in higher education, AAC&U's reports provide supporting evidence for each of the outcomes from business, government, faculty, and institutions. AAC&U also developed rubrics for 15 identified essential learning outcomes, making them much easier to operationalize and apply to campus practices (AAC&U, 2011). Additionally, the consensus in AAC&U's research among different sectors of the community and those earlier taxonomies and theories support the importance of the above higher education learning outcomes.

Inspired by the work of the AAC&U, researchers at Wabash College designed a study to investigate certain outcomes related to liberal arts. These outcomes focus more on ethical development, moral character, leadership, and inquisitive minds than other areas of the general outcomes. The Wabash National Study of Liberal Arts Education (hereafter, the Wabash study) is a large-scale, longitudinal study that aims to explore the institutional conditions and practices that improve liberal arts education outcomes (Center of Inquiry, 2009). Resulting from two foci of connecting the "qualities of mind" and the "responsibilities of citizenship" in the process of promoting holistic student development, the research team identified seven specific outcomes: (1) integration of learning, (2) inclination to inquire and lifelong learning, (3) effective reasoning and problem solving, (4) moral character, (5) intercultural effectiveness, (6) leadership, and (7) wellbeing (King, Kendall Brown, Lindsay, & VanHecke, 2007).

Some projects have tried to define general education outcomes under a transformative paradigm and an integrated point of view. Learning Reconsidered – a report jointly published by two higher educational associations: The National Association

of Student Personnel Administrators (NASPA) and The American College Personnel Association (ACPA) – categorized student learning outcomes into seven broad areas (NASPA & ACPA, 2004). These outcomes include: cognitive complexity; knowledge acquisition, integration, and application; humanitarianism; civic engagement; interpersonal and intrapersonal competence; practical competence; and persistence and academic achievement. Also derived from the perspective of transformational learning, the paper "A Well-Rounded Education for a Flat World", prepared by the College Outcomes Project, elaborated the learning outcomes of a successful liberal education for "the flat world of 21st century" (Hersh et al., 2008). They envision transformational learning outcomes under two broad categories:

> The first category, the "traditional," primarily cognitive outcomes of college, includes knowledge and reasoning in the disciplines, as well as the broad abilities of critical thinking, analytic reasoning, problem solving, and communication. The second category contains those outcomes addressing psychosocial, affective, and interpersonal dimensions, inextricably connected, to cognitive development. (p.10-11)

Although they further developed sub-categories under each category, these two broad categories are probably the most parsimonious classification available. This categorization is consistent with Astin's two types of outcomes – cognitive and noncognitive, which behavioral sciences have used for decades. Additionally, the College Outcomes Project clearly stated that they did not try to exhaust all learning

outcomes by listing the categories. Rather, they aimed to provide an integrative list of outcomes of importance.

In a series of papers titled *A Culture of Evidence*, scholars at ETS summarized three major dimensions of student learning at the postsecondary level – workforce readiness and general education skills, content knowledge/discipline-specific knowledge and skills, and "soft skills" (noncognitive skills). They also provided a fourth dimension – student engagement[3] – stating that it is "important to student success and should be carefully monitored" (Dwyer et al., 2006, p.13).

In a recently released report, *The Degree Qualifications Profile*, the Lumina Foundation for Education "propos[ed] specific learning outcomes that benchmark the associate, bachelor's and master's degrees… regardless of a student's field of specialization" (Adelman et al., 2011, p.1). Five primary areas of competence were set forth – specialized knowledge, broad/integrated knowledge, applied learning, intellectual skills and civic learning. In their structured and interconnected spiderweb, these five competencies formed five anchor lines which extend from the center of the learning outcome web to different levels of degree qualifications (Adelman et al., 2011).

The various classifications of higher education learning outcomes introduced in this section are only a sample of different classifications. Many other national associations and researchers have made similar attempts to achieve the same goal; due to space considerations, they are not all presented here (see American College Personnel Association, 1994; Baxter Magolda, 1999; Business-Higher Education Forum, 2003;

[3] They did not include student engagement as one of the major dimensions because it is "not in itself a student learning outcome" (Dwyer et al., 2006, p.13).

Immerwahr, 2000; Western Association of Schools and Colleges, 2008). An interesting finding is that albeit having the same goal, none of the two classifications are exactly the same, which highlights one aspect of complexity involved in the learning outcome literature. This is one of the reasons why this study uses a more simplified classification system which is introduced in the next section. It also discusses complex conceptualizations of important constructs and clarifies their meanings.

2.1.3 A Clarification on the Concepts

From Astin's (1973) seminal work on a higher education outcome taxonomy and student development theories to various more recent collaborations among colleges and universities, higher educational associations, accrediting organizations, and business communities, consensus has long existed among researchers that student learning in higher education has multiple dimensions. Although classification of these dimensions varies in different researchers' work, two major constructs are consistent throughout.

One construct involves students' cognitive development, which can include general education knowledge and skills, discipline-specific knowledge, and practical skills for the workforce. Because of the longstanding tradition of focusing on students' verbal and quantitative knowledge measured by standardized tests, this construct has been widely accepted and used. The second construct involves students' noncognitive development, such as attitudes and beliefs, values, motivation, and other psychosocial factors. This construct is broad, yet has received little attention in education literature.

Therefore, it is not surprising that conceptual confusions arise when looking at this construct and its multidimensional components.

As Robbins et al. (2004) stated in a meta-analysis study about college outcomes, "surprisingly, there is little integration or research synthesis of the educational and psychological literatures when looking at college outcomes" (p.261). Confusion occurs in these two fields of literature even when defining the two basic college outcomes. Within the education literature, cognitive outcomes have long been used to refer to "standardized academic achievement and aptitude tests and school-based academic performance (e.g., grade point average [GPA] and class rank)" and noncognitive can refer to all the remaining outcomes (p.261). However, within the literature of cognitive psychology, "a broad range of constructs are viewed as cognitive, including self-concepts such as self-efficacy beliefs and outcome expectancies, meta-cognitive knowledge, and achievement and performance goals" (p.261). The aforementioned cognitive theories that focus on "*how* people think, rather than just *what* people think" serve as a good example of the cognitive construct in the psychological literature. Another example for the education literature is the use of the noncognitive concept in a series of papers by ETS. In their research, "soft-skills" such as creativity, teamwork and persistence were put under noncognitive skills category and were separated from domain specific and general education skills (Dwyer et al., 2006).

Unfortunately, the distinction between these two systems of definitions is not as clear cut across different studies. Researchers in different disciplines may use the same term referring to the same construct, while those within the same discipline may use

different terms interchangeably. Acknowledging this conceptual confusion, some researchers simply abandon the use of *noncognitive* and replace it with other terms, such as psychosocial (Martin, Swartz-Kulstad, & Madson, 1999; Robbins et al., 2004), social emotional (Hoffman, 2009), nonacademic (Hearn, 1991), socioaffective (Neihart, 2007), affective-motivational (Cotic & Zuljan, 2009), personal qualities (Willingham, 1985), cognitive balance (Heider, 1958, as cited in Messick, 1979), personal constructs (Kelly, 1955, as cited in Messick, 1979) and attributions (Kelley, 1973 and Weiner, 1972, as cited in Messick, 1979). While acknowledging the label *noncognitive* as a misnomer (Kyllonen, 2005; Messick, 1979), most researchers adopt the noncognitive term, because of its wide use in industry, psychology, and measurement (Kyllonen et al., 2005; Messick, 1979; Sedlacek, 2004).

Recognizing the conceptual confusion, this dissertation uses the widely accepted cognitive and noncognitive terminologies from the education literature. The term cognitive is used to refer to student academic-related achievement which is typically measured by standardized tests, such as general education knowledge and skills, discipline-specific knowledge, and practical skills, and sometimes demonstrated through coursework, such as writing samples, specialized projects, and senior theses. The term noncognitive is used to refer to everything else, especially those constructs related to students' personality, attitudes and beliefs, values, and motivations. Other terms, such as psychosocial, if used, refer to other researchers' work and are interchangeable with the noncognitive construct. To avoid potential confusion, the other terms are not used beyond this chapter.

2.2 Noncognitive Constructs & Civic Engagement

2.2.1 The Importance of Noncognitive Constructs

Postsecondary education has focused on students' intellectual development and vocational training, especially in the twentieth century. Standardized academic achievement tests have been used and are playing an important role in higher education admission and in predicting success in education. Accordingly, research in higher education has focused almost exclusively on students' development of cognitive abilities while ignoring noncognitive constructs.

However, in the field of industrial and organizational psychology, noncognitive constructs are assessed and valued. Noncognitive assessments are used by employers to make hiring decisions and evaluate job performances and training success (Colquitt, LePine, & Noe, 2000; Roberts & Hogan, 2001). Intellectual capacity is only a small part of the much broader concept of workplace competence (Hough, 2001). Noncognitive components that are related to social and emotional intelligence, such as integrity, motivation, citizenship performance, emotional stability, self-esteem, and locus of control, are important criteria in personnel practices (Colquitt et al., 2000; Schmidt & Hunter, 1998).

After the series of reports from AAC&U, more researchers and stakeholders in the field of education are starting to acknowledge the importance of noncognitive constructs both conceptually and empirically. From the conceptual aspect, many faculty members, administrators, and employers believe students' noncognitive attributes are strongly associated with their success at school and in the workplace. In an interview

study with graduate school staff and faculty about both admission and measurement of student success related issues, a number of noncognitive constructs were rated highly – including values and character (such as integrity, openness, honesty, and consistency), maturity and responsibility, interpersonal skills, commitment to the field, and willingness to take a challenge (Walpole, Burton, Kanyi, & Jackenthal, 2002).

AAC&U led a study in late 2009 to examine employers' views of college learning. Over 300 employers with at least a quarter of new hires holding either an associate's degree or bachelor's degree and with a minimum of 25 employees were interviewed. Among the respondents were CEOs, presidents, vice presidents, owners, and other top executives at private sector and non-profit organizations. Seventy-five percent of the respondents value "the ability to connect choices and actions to ethical decisions"; 71% value "teamwork skills and the ability to collaborate with others in diverse group settings"; and 52% value "civic knowledge, civic participation, and community engagement" (Hart Research Associates, 2010). Additionally, this study largely supports findings from a 2007 AAC&U national survey study where employers repeatedly stressed the importance of providing all students with a well-rounded education with an emphasis on noncognitive dimensions such as teamwork skills, effective communication in diverse settings, a sense of integrity and ethics, and civic knowledge, civic participation, and community engagement (AAC&U, 2007).

Researchers have also collected ample evidence showing the importance of noncognitive attributes in education. Not only are there numerous findings in early childhood (e.g., Abe, 2005; Caprara, Barbanelli, Pastorelli, Bandura, & Zimbardo, 2000)

and K-12 areas (e.g., Artelt, Baumert, Julius-McElvany, & Peschar, 2003; Duckworth & Seligman, 2005), noncognitive attributes are found to be important during the four college years, at the graduate level, and after graduation (e.g., Heckman & Rubinstein, 2001; Robbins, Allen, Casillas, Peterson, & Le, 2006; Sedlacek, 2004). Based on the multiple roles noncognitive factors play in the higher education literature, the next section is broken into two parts – noncognitive constructs serving as predictors and as outcomes.

2.2.1.1 Noncognitive constructs serving as predictors

Noncognitive measures are not often used to assess students, especially in higher education. However, among the studies that treated noncognitive factors as predictor variables, factors such as motivation, social involvement, and self-management are shown to be important in predicting postsecondary students' academic performance and persistence (Chamorro-Premuzic & Furnham, 2003; Robbins et al., 2004; Robbins et al., 2006; Sedlacek, 2004; Wagerman & Funder, 2007). In a meta-analysis, Robbins et al. (2004) found moderate relationships between college retention and academic goals and self-efficacy. As for students' college GPA, one of the strongest predictors found was motivation.

Sedlacek (2004) also argues for the importance of noncognitive constructs. He has designed the Noncognitive Questionnaire (NCQ) to assess attributes that are related to success in higher education such as positive self-concept or confidence, realistic self-appraisal, successful leadership experience, and demonstrated community service. Over

much of the past three decades, he and his colleagues found positive associations between noncognitive constructs and college achievement and retention (see Boyer & Sedlacek, 1988; Sedlacek, 2005; Tracey & Sedlacek, 1984, 1988). Moreover, these noncognitive variables are not only useful predictors for traditional college students, but are critical for nontraditional students whom he described as a large group of students including women, racial/ethnic minority groups, student athletes, older students and all others "with cultural experiences different from those of White, middle-class, heterosexual, males of European descent" (Sedlacek, 2004, p.4).

Considerable amounts of variance in academic achievement and attainment are accounted for by noncognitive factors beyond traditional predictors such as high school grades and SAT® scores, even when freshman GPA is included (Robbins et al., 2004; Robbins et al., 2006; Wagerman & Funder, 2007). Robbins et al. (2006) built models with four blocks of independent variables to predict college retention, cumulative GPA, and grades on English and mathematics courses. Although entered into models in the last block, psychosocial factors were found to have incremental predictive validity after controlling for institutional characteristics, student demographics, and measures on prior academic achievement such as high school GPA and ACT scores. Specifically, in the hierarchical linear regression models predicting first-year GPA in four-year institutions, the psychosocial factors explained an additional 3.4% of the total variance in GPA, with an R^2 change from 0.356 in the model without psychosocial factors to 0.390 in the final model including psychosocial factors. Moreover, in the hierarchical logistic regression models predicting student first-year retention and course successes on English

composition and algebra (here, course success is defined as having a grade of B or higher), the overall odds ratio increased from 1.92 to 2.06, from 2.75 to 2.90, and from 3.22 to 3.42, respectively.

Furthermore, the importance of noncognitive skills continues years after graduation from college, which is evident in training and workforce performance (Heckman & Rubinstein, 2001; Schmidt & Hunter, 1998). Some researchers consider job performance a critical part of higher education outcomes and have used noncognitive factors to predict student success after graduation. Several meta-analyses examined and confirmed that conscientiousness, as well as emotional stability and extraversion, are strong predictors of job performance (Barrick & Mount, 1991; Hurtz & Donovan, 2000; Salgado, 1998; Tett, Jackson, & Rothstein, 1991).

2.2.1.2 Noncognitive constructs serving as outcomes

Pascarella and Terenzini (1991, 2005) synthesized large amounts of literature from 1967 to 2002 that looked at college's effects on students' psychosocial changes, their development of attitudes and values, and on moral development. In most of the studies reviewed, noncognitive factors serve as outcome variables and college attendance or year in college serves as one of the predictors, although different studies vary in their measures of noncognitive constructs and in the other variables controlled for. As Pascarella and Terenzini pointed out, Knox, Lindsay, and Kolb (1993) provided "perhaps the most rigorous evidence of the influence of college on students' involvement in civic and community affairs" (p.289). This statement still holds true today. Using longitudinal

and nationally representative data, Knox, Lindsay, and Kolb's (1993) study found a statistically significant but small effect from college attendance on students' locus of control, after controlling for student characteristics and academic ability. In other studies, college attendance has also been found to have strong associations with students' multicultural attitudes and values (Case & Greeley, 1990), gender-role attitudes (Astin, 1993a; Miller-Bernal, 1993), occupational values (Knox et al., 1993), and moral reasoning (Wilson, Rest, Boldizar, & Deemer, 1992). More specifically, attending a liberal arts college has been demonstrated to have an impact on openness to diversity and challenges and learning for self-understanding (Pascarella, Wolniak, Serfert, Cruce, & Blaich, 2005).

In *What Matters in College*, Astin (1993a) used his renowned conceptual framework for studying student outcomes, the input-environment-outcome (I-E-O) model, to extensively examine the effect of college on various noncognitive constructs including personality and self-concept, attitudes, values, and beliefs. In the I-E-O model, inputs refer to student characteristics before entering the institution; environment refers to various college experiences the student is exposed to; and outcomes refer to the student's characteristics after spending time in college (Astin, 1970, 1993a). Two basic ideas behind this model are worth mentioning. One is that environmental experiences are critical to students' changes during college. In other words, institutional culture and students' engagement in a wide variety of campus activities are important factors to take into consideration in predicting postsecondary outcomes. From the administrator's point of view, student development on the postsecondary outcomes can be maximized by

structuring a supportive and diverse educational environment. The other is that the observed correlation between the outcome(s) and the environmental variable(s) cannot be properly assessed until the input variables are controlled for. These input variables can be student background characteristics such as gender, race, and family background. They can also be some kind of pre-measure of the outcome prior to any exposure to the institution. This latter group of input variables is a basic component in models that measure value-added, which is discussed in more detail later.

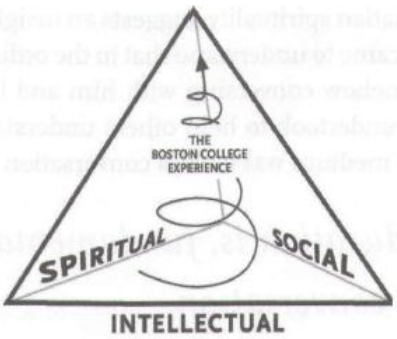
Astin's (1993a) extensive study found associations between institution-level variables and some noncognitive constructs. After controlling for student characteristics and pre-college self-concepts or corresponding pre-college values and attitudes, most variables at the institution-level (such as institutional type and control) were not found to be associated with students' changes in attitudes and values anymore. However, a number of institutional environmental and involvement factors were found to be associated with development of attitudes and values during the college years. For instance, student peer group experiences (e.g., frequent student-faculty and student-student interactions, faculty's values and attitudes) appeared to have a significant effect on social activism, status striving, as well as students' views on social issues. The effect of peer involvement on measures of self-concept and psychological well-being was found to vary among different forms of involvement in college, while the effect of diversity experiences had a significant positive impact on students' values such as liberalism, feminism, commitment to promoting racial understanding, and developing a meaningful philosophy of life.

In this new century, colleges and universities, policy makers, researchers, as well as parents and students themselves have more clearly recognized the multiple roles noncognitive skills play. Some have suggested that noncognitive factors be added to admission criteria for both undergraduate and graduate students' enrollment (see Kyllonen et al., 2005; Sedlacek, 2004), used in career guidance or self-diagnosis such as helping students select an undergraduate major or better understand themselves (Kyllonen, 2005), and used to identify at-risk students who need specific intervention from the institution (Robbins et al., 2006).

In practice, more institutions are starting to realize the importance of facilitating student development with respect to noncognitive attributes beyond academic learning – developing students' social responsibility, "learning about others and oneself," promoting positive personal attitudes and values, fostering transferable intellectual and practical skills to real-world settings, and preparing for a democratic society and a globalized economy (Beyer, Gillmore, & Fisher, 2007; Boston College, 2007; Indiana University, 2010; Lewis, 1997; University of Michigan, 2011; University of Wisconsin System, 2010).

To achieve the goal of fostering well-rounded citizens, many higher education institutions establish comprehensive student development strategies. Presented in *The Journey into Adulthood* (Boston College, 2007), Boston College has conceptualized an intellectual, social, and spiritual three-dimensional pyramid (see Figure 2.1) to represent the integrated identity the college seeks to develop in each student over the four years of undergraduate education. The focus is on helping students to "develop their gifts of mind

and heart, cultivate their interior lives, and make good decisions about how they will use their gifts to help others" (p.9). Shaped specifically through the social and societal dimension of the student formation pyramid, Boston College fosters students' experiences in "learning to live with people different from oneself," "developing social skills and leadership abilities," and enlarging "their understanding of who is to be cared for and who is to be included in their community of concern [across racial, ethnic, cultural, and socio-economic boundaries]" (p.10).



Source: Boston College, 2007, p.13

Figure 2.1 One Process Moving to Integration

As seen in the earlier section on Dimensions of Student Learning Outcomes, there is widespread agreement on the vision of improving students' civic responsibility and engagement among higher education researchers and educators. The next section highlights the current understanding of civic-related issues.

2.2.2 Civic Engagement

2.2.2.1 Operational definition

As civic engagement has been used in various studies with a variety of meanings, an operational definition will help to clarify what civic engagement means in this study. Based on Ehrich's (2000) definition, the AAC&U extended the meaning of civic engagement in the rubric of Valid Assessment of Learning in Undergraduate Education (AAC&U, 2010):

> Civic engagement is "working to make a difference in the civic life of our communities and developing the combination of knowledge, skills, values and motivation to make that difference. It means promoting the quality of life in a community, through both political and non-political processes." (Excerpted from Civic Responsibility and Higher Education, edited by Thomas Ehrlich, published by Oryx Press, 2000, Preface, page vi.) In addition, civic engagement encompasses actions wherein individuals participate in activities of personal and public concern that are both individually life enriching and socially beneficial to the community. (p.1)

This definition encompasses both inward and outward aspects of people's propensity toward civic engagement for the public good. In other words, as stated in the benchmark descriptions in the AAC&U rubric, civic engagement can be demonstrated broadly, including participation in various civic activities, demonstration of the ability of connecting and extending one's own knowledge to civic actions, as well as the

commitment and adjustment in attitudes and beliefs towards civic involvement, and the "reflective insights or analysis about the aims and accomplishments of one's actions" (AAC&U, 2010, p.2). This is evident in the various names researchers have given to this corresponding construct – "civic responsibility" (Colby et al., 2003; Myers-Lipton, 1998), "civic orientation" (Crystal & DeBell, 2002), "citizenship" (Sax, 2000; Smith, Nowacek, & Bernstein, 2010), "civic and community attitudes and involvement" (Pascarella & Terenzini, 2005), "civic values" (Rhee & Dey, 1996), "civic virtue" (Sax & Astin, 1998), "humanitarian/civic involvement values" (Pascarella, Ethington, & Smart, 1988), "social action engagement" (Nelson Laird, Engberg, & Hurtado, 2005), or "other orientation" (Villalpando, 1996), among others. Since the term "civic engagement" has been widely used and accepted across a broad range of stakeholders, in the interest of being concise, "civic engagement" is used in this study to refer to both the civic activities students participate in and their attitudes and values towards these civic activities.

2.2.2.2 Empirical findings

The effect of college on students' civic engagement has gained increasing interest since the 1990s and received a comprehensive review in Pascarella and Terenzini's (2005) book. Similar to what has been found in noncognitive outcomes in general, the authors found a consistent positive relationship between commitment to civic life and community activities and college attendance. Moreover, college's effects on students' civic and community attitudes and involvement were found to "endure well beyond graduation" (p.330).

It is worth noting that different conceptualizations and different measures are used in different studies examining civic engagement. In these studies, civic engagement can involve one or multiple components, including commitments to and involvement in community service, volunteer work, community leadership, voting, other political activities, racial understanding, and helping others who are experiencing difficulty (see Berkner, He, & Cataldi, 2002; Rhee & Dey, 1996). There are also differences in how the construct is measured and how the indicator of civic engagement is derived. For example, questions are written differently in different surveys; different combinations of the questions can be selected; and different composition approaches can be used to form the final civic engagement outcome. Even in the studies that used similar CIRP survey items to derive civic engagement related measures, different methods were applied. Some studies used a single item to represent one aspect of civic engagement; some created composite scores by adding multiple civic related items; some others calculated factor scores based on results from exploratory factor analysis. Therefore, we should keep in mind the different conceptualizations and the different measures may result in inconsistent findings with respect to civic engagement.

Using national representative data, a number of studies have found a consistent positive association between education level and civic and community activity involvement (Berkner et al., 2002; Kuh et al., 2001; Nolin, Chaney, & Chapman, 1997). For instance, the Center for Information and Research on Civic Learning and Engagement (CIRCLE) examined data from the Current Population Survey and the National Civic Engagement Survey among 18 to 25 year old American youths. Results

from simple frequency analysis showed that those who received some college education scored higher across measures including civic involvement, electoral participation, and political voice compared against peers with no college experience (Lopez & Elrod, 2005, 2006; Lopez & Brown, 2006).

At the same time, these cross-sectional findings also found evidence of support from nationally representative, longitudinal studies (Huang & Healy, 1997; Ingels, Curtin, Kaufman, Alt, & Chen, 2002; Knox et al., 1993; Sax, 2000; Sax & Astin, 1998; Villalpando, 1996). While examining a four-year longitudinal database from the CIRP project, Sax and Astin (1998) compared students' citizenship qualities between their freshman and senior year of college. They summarized the effect of various college environments and student experiences on the improvement on their sense of civic virtue, and demonstrated higher education's ability in citizenship development. Using longitudinal data from the National Education Longitudinal Study of 1988, Ingels and his colleagues concluded that the likelihood of volunteering among those with reported bachelor's or higher degrees are about 2.5 times higher than those without postsecondary experience (Ingels et al., 2002).

Moreover, using data from the CIRP freshman survey and the follow-up survey administered in 1971 and 1980, Pascarella et al. (1988) found significant associations between social leadership involvement with peers during college (such as "president of one or more student organizations" and "served on a university or departmental committee"), post-college social service occupation (such as social work and counselor), and respondents' civic engagement values. Similarly, Astin and Sax (1998) found

positive correlations between different types of community service participation (e.g., education, human needs, public safety, or environment), time devoted to community service, and students' sense of civic responsibility.

In general, college attendance has shown a positive impact on students' civic engagement, even when controlling for precollege differences (Pascarella & Terenzini, 2005). Knox et al. (1993) found a clear and consistent association between college attainment and civic involvement after controlling for students' background characteristics, such as gender, race/ethnicity, socioeconomic status, and their pre-college civic attitudes and values. Other than college attainment, many other researchers explored the effect of various college experiences on students' development of civic engagement after controlling for students' precollege characteristics, institutional characteristics, or college activities. These experiences include social involvement during college such as social leadership involvement with peers, familiarity with faculty/staff, having a social science major (Pascarella et al., 1988), discussing racial/ethnic issues or political/social issues with peers, participating in campus demonstrations, socializing with someone of a different ethnic group (Rhee & Dey, 1996; Sax & Astin, 1998), and participating in community service activities (Astin, 1993a; Sax & Astin, 1997; Villalpando, 1996; Astin & Sax, 1998; Gray et al., 1999). Results from a more recent meta-analysis are also consistent with prior findings which showed a consistently strong relationship between college diversity experiences and students' development in civic attitudes, behavioral intentions, and behaviors (Bowman, 2011).

After taking into account student characteristics, variables often used to classify college quality and to distinguish differences among institutions, however, are found to have little or no impact on changes in students' civic engagement. A handful of studies have identified a few exceptions for certain institution-level variables. For instance, the National Survey of Student Engagement (2004) project concluded in their annual report that "students at historically Black colleges and universities are far more likely to participate in a community project linked to a course and report gaining more in personal, social, and ethical development" (p.12). Rhee and Dey (1996) found little indication that attending religiously affiliated colleges, as opposed to their nonsectarian counterparts, has a positive and direct effect on students' attitudes towards civic responsibility, volunteerism, and civic and social values. Some have found that students' civic values are negatively correlated with institutional selectivity and, sometimes, institutional size (Astin, 1996; Pascarella et al., 1988). However, this negative correlation may be a result of activities encouraged at the particular campuses (Astin, 1996). Moreover, using path analysis, Smith, Wolf, and Morrison (1995) found that attendance at a women's college is an indirect predictor of students' civic involvement during college, mediated by institutional priorities and students' extracurricular involvement.

Various studies have shown the development of civic engagement during postsecondary study. However, it is worth comparing students' development in this area with other education outcomes. Based on student self-reported data in 2004, NSSE revealed senior students' reported gains in different education outcomes. While 86% of seniors reported that their college contributed "quite a bit" or "very much" to them

acquiring a broad general education, only 45% of them said their college contributed to their development of a concern for the welfare of their community and 23% in voting in local, state, or national elections. In the "civic responsibility and engagement" area, student gains are often less than hoped for (AAC&U, 2005). From another perspective, the comparison between students' development in civic engagement and other areas confirmed employers' concern about prospective employees being well-rounded citizens and receiving holistic educations.

2.2.2.3 Civic engagement in the framework of higher education outcomes

The above overview summarized the findings regarding noncognitive outcomes, and civic engagement specifically, in higher education and beyond. Since noncognitive outcomes are multidimensional constructs and involve many components that may change during college, how is civic engagement viewed in the current framework of higher education outcomes? To highlight the importance of civic engagement in the overall framework, the table below summarizes the consensus on this issue from the perspectives of different national higher education associations, regional and specialized accreditors, institutional research centers, foundations, and other research scholars.

Table 2.1 Perspectives of Civic Engagement as Part of Higher Education Outcomes

Research conducted by	Civic engagement related higher education outcomes
National Higher Education Associations	
Association of American Colleges & Universities	Personal and social responsibility, including civic knowledge and engagement – local and global
American College Personnel Association	A coherent integrated sense of identity, self-esteem, confidence, integrity, aesthetic sensibilities, and civic responsibility
The National Association of Student Personnel Administrators & The American College Personnel Association	Civic engagement: sense of civic responsibility; commitment to public life through communities of practice; engage in principled dissent; effective in leadership
The National Postsecondary Education Cooperative (NPEC) Working Group on Student Outcomes from a Policy Perspective	Civic development: Group affiliations/memberships, citizenship, community involvement, voting participation
Regional and Specialized Accreditors	
The Association to Advance Collegiate Schools of Business	Knowledge and skills in ethical and legal responsibilities in organizations and society; Ethical understanding and reasoning abilities; Multicultural and diversity understanding
Accreditation Board for Engineering & Technology	Knowledge of contemporary issues; Understanding professional and ethical responsibility
Accrediting Commission for Senior Colleges & Universities, Western Association of Schools & Colleges	Appreciating civic responsibility; Appreciating diversity
Western Association of Schools & Colleges	An understanding of diversity; civic responsibility
Institutional Research Centers	
The Center of Inquiry in the Liberal Arts at Wabash College	Leadership, including group values and a societal and community value (citizenship); Moral character, involves the capacity to make and act on moral or ethical judgments, treating others with fairness and compassion
Foundations	
Charles Engelhard Foundation – The College Outcomes Project	Perspective-taking/acting, directly connected to civic engagement, ethical and moral reasoning/action, and social-culture competence
Lumina Foundation for Education	Civic learning

Table 2.1 (cont.)

Research conducted by	Civic engagement related higher education outcomes
Other Research Scholars	
Baxter Magolda (1999)	Interpersonal competence, involves humanitarianism and concern for the community
Oswald, Schmitt, Kim, Ramsay, and Gillespie (2004)	Interpersonal behaviors – Multicultural, Leadership, Interpersonal, and Citizenship

Resources: Adelman et al., 2011; AAC&U, 2004, 2005, 2007; ACPA, 1994; Baxter Magolda, 1999; Hersh et al., 2008; King, Kendall Brown, Lindsay, & Vanhecke, 2007; NASPA & NCPA, 2004; Oswald et al., 2004; U.S. Department of Education, NCES, 1997; and Western Association of Schools and Colleges, 2008.

2.3 Current Outcome Measurements and Challenges

Unlike the K-12 setting, there has been limited national focus on the evidence of student outcomes in higher education. Recent concerns on the part of government and the general public about higher education institutional quality and competency in the 21st century are turning the emphasis to one of accountability. Many institutions have started to develop and use local assessments to measure student outcomes and evaluate individual programs. However, to achieve the goal of reporting comparable and meaningful student outcomes for the purpose of accountability, only nationally or regionally administered outcome measures that allow us to compare student outcomes across different institutions are the focus of this section.

Previous research organizes four general approaches to collect data and assess higher education quality – actuarial data, ratings of institutional quality, student surveys, and direct measures of student learning (Chun, 2002). Although several improvements in measuring higher education outcomes have been made in the past decade, these four approaches are still applicable to today's assessment framework in higher education.

Therefore, each of these four approaches are reviewed in this section; then the challenges in measuring higher education outcomes are presented.

2.3.1 Outcome Measures in Accreditation

Evolving for over 100 years, higher education accreditation in the U.S. has been "the primary means by which colleges, universities and programs assure quality to students and the public" (Eaton, 2009, p.2). Accountability is widely used to ensure the performance of higher education institutions meets minimum standards in accreditation. Among most professional organizations in higher education, accrediting agencies are the sole bodies responsible for accountability and quality of colleges and universities other than those institutions themselves (Shavelson, 2010). Therefore, the first to be examined is accreditation – the processes and the outcome measures used by accrediting agencies.

Accreditation in U.S. higher education is "a collegial process… for improvement of academic quality and public accountability of institutions and programs" (CHEA, 2006, p.2). National, regional, and programmatic accrediting organizations normally carry out their quality review processes through a series of steps: (1) Self-study: Based on the accrediting agency's standards, institutions or programs self-develop a summary of descriptions which normally includes goals, missions, governance, faculty, curriculum, facilities, and self-evaluation; (2) Peer review: "Faculty and administrative peers in the profession" review the self-study results and serve as board members to make judgments on accrediting status; (3) Site visits: A visiting team is sent to review the institution or the program which provides feedback in addition to the self-study and peer review; (4)

Decision or judgment: Based on the available information, accrediting agencies make decisions to grant, reaffirm, or deny accreditation to institutions and programs; and (5) Periodic external review: Institutions and programs will be reviewed periodically over time (Eaton, 2009, p.4).

One would assume that the national, regional, and programmatic accrediting organizations have collected large amounts of useful information for the purpose of institutional accountability. However, traditionally, accreditation only examines inputs and processes such as student-faculty ratios, capital resources like libraries and classrooms, professors' credentials, program review procedures, and whether the institution has a mission statement and a faculty senate. Not many concrete measures or objective data on student learning outcomes are examined by accrediting organizations. Moreover, as the Spellings Commission and many critics have pointed out, accreditation has "significant shortcomings… Accreditation reviews are typically kept private" (U.S. Department of Education, 2006, p.14). In other words, the limited accreditation information is only used internally by colleges and universities. What the general public learns is whether an institution is accredited or not. The public's need for transparent and reliable information about higher education institutions is not being met (American Council of Trustees and Alumni, 2007; Dickeson, 2006; Lederman & Redden, 2007; Leef & Burris, 2002; Malandra, 2008).

More recently, the review processes have often involved some form of assessment; however, the mode of assessment often has varied greatly. Since most of the evidence is provided by individual institutions based on internal self-studies, and institutions are

required to include their own list of intended outcomes linked to their specific education goal and mission, different assessment results are often provided across campuses, making it difficult to make comparisons. Moreover, those programmatic or specialized accreditors adopt particular requirements for specific programs and professional schools such as the medicine, law, engineering and technology professions. Therefore, the collected information is not sufficient to answer the questions, "Compared to what? Compared to whom?" (conversation with Margaret Miller, as cited in Lederman & Redden, 2007). From a research point of view, different standards applied and uneven information provided across different colleges and universities lack consistency and make it nearly impossible to compare nationwide or regionally.

Facing these challenges and the threat of the loss of credibility, the accreditation organizations have decided to "model the value of outcomes measurement" (Dickeson, 2009, p.18). Although accreditation has not been able to elicit much useful data that provide outcome measures in higher education, currently, changes are being made on emphasizing measuring college outcomes. In the future, it is hoped that the accreditation organizations can collect valuable information to provide comparable evidence of student outcomes and the effectiveness of institutions and programs.

2.3.2 Ratings and Rankings

Another approach to assess higher education quality is to have college faculty members and administrators rate different institutions and programs. The well-known college rankings published by *U.S. News & World Report* also uses this as one of their

ranking criteria. While there are other ranking systems available, such as those made by the Princeton Review and Forbes, since the *U.S. News* rankings is the most well-known and receives the most attention, their ranking measures are briefly introduced here.

According to the director of data research at *U.S. News*, Robert Morse, the 2011 rankings are based on key measures in seven categories: undergraduate academic reputation, which is calculated based on ratings from administrators at peer institutions and high school counselors; student selectivity; faculty resources; graduation and retention rates; financial resources; alumni giving; and graduation rate performance (Morse, 2010). Numerous critiques have emerged about the methodologies *U.S. News* adopts, including the weighting scheme, face validity of the measures, credibility of the reputation ratings, and subjectivity of the measures included, as well as their impact on institutions and students, and the role *U.S. News* plays in the accountability system (e.g., Clarke, 2002; Machung, 1998; Monks & Ehrenberg, 1999; National Opinion Research Center, 1997; Thompson, 2000). Moreover, the recent Claremont McKenna College scandal, where Claremont McKenna admitted reporting inflated SAT scores since 2005 to be more competitive in rankings of elite liberal arts colleges, illustrates concerns about credibility and data accuracy underlying the ranking system (Slotnik & Pérez-peña, 2012).

For the purpose of this dissertation, the heart of the problem in the ranking system is that none of the measures used in the *U.S. News* rankings are related to student learning outcomes during college. In other words, the ranking criteria predominantly focus on input variables, such as student selectivity, capital resources, and endowment giving, rather than learning outcomes, learning environment, or how much students gained from

their college experience. The primary outcomes considered are the average graduation and retention rates of the institution and graduation rate performance which is calculated as the difference between the actual and expected graduation rates. The weight of these outcome measures is only 27.5% of the total. To date, the literature has found no evidence that the ranking is directly linked to the actual student learning outcomes, which is the most crucial and direct measure in estimating school effectiveness. As Mr. Morse agrees, "I wish that we were able to measure things like outcomes in learning…[But] there just isn't any data available. The schools themselves aren't measuring learning" (Fitzpatrick, 2009). In truth, many schools are measuring learning outcomes, but not in ways that can be compared across schools as *U.S. News* would need.

Strictly speaking, ratings and rankings of institutional qualities do not involve outcome measures of what students have learned during their college years. The subjective nature of the *U.S. News* ranking system and accreditation systems perpetuate similar shortcomings. The ranking system's advantage, however, is to provide publically available information to all stakeholders. From another perspective, it is not hard to understand the popularity of the *U.S. News* annual college rankings and their self-appraised "#1 comprehensive" college guide book. It reflects the great demand among prospective students, their parents, and other higher education consumers to compare across institutions before making one of the largest and most influential investments of their lives. Regardless of whether they have heard of the controversial reputation of the *U.S. News* ranking, these consumers have no other choice because there are no better options, and they desire to make as informed of a decision as possible.

2.3.3 Direct Measures of Learning Outcomes

2.3.3.1 Before Voluntary System of Accountability

The above approaches so far have shown that without evidence of learning outcomes (i.e., what students have learned and can do), any method of assessing higher education quality is of limited utility and susceptible to challenge. Because of this, many researchers advocate measuring student learning outcomes directly (e.g., Callan & Finney, 2002; Klein, 2002; Shavelson, 2010). Such direct measures can adopt the format of standardized assessments (e.g., GRE, licensure tests, end-of-course tests), course grades, and student portfolios. Moreover, the content of these measures can include discipline-specific knowledge as well as broad abilities such as reasoning and general intelligence.

However, most of the currently available direct measures are locally developed and used at the classroom, program, or campus levels (Kuh & Ikenberry, 2009; Ratcliff, Jones, Guthrie, & Oehler, 1991). More specifically, course grades are associated with individual professors; judgments on portfolios and other classroom performances can easily raise validity concerns; and only a small portion of the students (those who need the test results to apply to graduate schools or need to obtain a license before working as a teacher, nurse, and etc.) take the GRE and licensure tests. As a result, these measures have not commonly been used to compare across institutions nationwide or regionally. Other than widely adopted institution-centered assessment approaches, some exceptions exist. For example, some institutions have conducted collaborative studies to compare student learning outcomes amongst themselves (Obler, Slark, & Umbdenstock, 1993). Some states mandate all institutions use the same standardized instruments to measure

common knowledge (Nettles, Cole, & Sharp, 1997; U.S. Department of Education, 1996). However, even when common statewide instruments are available, problems still exist in making comparisons across different instruments used in different states. As summarized by Chun (2002), "although it may seem to be the most obvious way to assess the quality of undergraduate education, the use of direct measures of student learning is uncommon" (p.25).

2.3.3.2 Voluntary System of Accountability

Given the traditionally decentralized higher education community, a collaborative effort is needed to form a nationwide data and accountability system. Responding to this need, in 2007 the VSA grew out of collaboration by two leading organizations in U.S. higher education – the AASCU and the NASULGC. The learning outcomes VSA currently measures are "common, multidisciplinary, and university-wide" cognitive outcomes including critical thinking, analytic reasoning, problem solving, and written communication (VSA, 2008a, p.2). VSA uses three standardized outcome instruments that each institution can choose from to measure these core education outcomes – the CAAP, the ETS Proficiency Profile, and the CLA. Although meaningful comparisons can be hindered, as each institution can choose one of the three standardized instruments, VSA is the best effort so far to provide direct college outcome information to prospective students, their parents, and other stakeholders.

For the purpose of this dissertation, the most problematic issue is the limited outcomes VSA measures. As can be seen from the broad missions of higher education

institutions, the purpose of higher education is not limited to educating students with respect to cognitive knowledge, but also cultivating well-rounded citizens with positive social responsibilities and advancing open-minded youths that serve society. As student learning outcomes in higher education have multiple dimensions, so must the instruments that can be used to measure them. Unfortunately, the three instruments VSA selected only measure students' cognitive outcomes which are defined as skills in written communications, critical thinking, and analytic reasoning. Students' noncognitive outcomes are not measured or used as part of the value-added measure in the VSA.

The other issue worth noting is the value-added methodology VSA adopts. Based on the value-added methodology initially employed for CLA, all three selected standardized assessments in VSA currently utilize a cross-sectional design. In order to isolate the approximate contributions of individual institutions (i.e., the value-added) and compare the progress their students make between freshman and senior years, two value-added approaches are generally used. One approach is cross-sectional, which tests freshman and senior students attending school at the same time. In this way, two different groups of students are tested once, usually at the same point in time. The difference score between the residuals when predicting each group of students' test scores with their college admission test scores is then used to estimate school value-added. The other approach is longitudinal, which tests students in their freshman year and tests them again in their senior year. In this way, the same group of students is tested twice. The difference between the same student's two test scores is the basis for calculating the value-added scores.

All the VSA assessments use a cross-sectional design because of the practical consideration that "the cross-sectional method is quicker, simpler, and less costly to implement" (VSA, 2008a, p.4). This cross-sectional methodology, however, is the cause of a major critique of VSA's measure of institutional value-added. For example, Garcia (2007) argues that senior students are "survivors" of the undergraduate education, so that they are a more selective population and not entirely comparable with the population of freshman students. Although VSA claims that there is no clear evidence to show the superiority of the longitudinal method over the cross-sectional method in calculating value-added scores (VSA, 2008a), in the large research field of both K-12 and higher education, the argument between the cross-sectional and longitudinal methodologies seems to have been settled in favor of the latter (more discussion is in section 2.3.6).

Under VSA's cross-sectional design, it is worth noting that the three testing agencies (i.e., ACT, CAE, and ETS) are currently using two value-added estimation methods. Since one of the foci of this dissertation study is value-added methodology, these two methods are presented in the following paragraphs. The first method is based on ordinary least squares regressions. It is the recommended method at the inception of VSA and used by all three testing agencies until the 2009-2010 assessment cycle. Starting from 2009-2010, ACT and ETS continue to use this method but CAE decided to use a second method which is based on hierarchical linear models (HLMs; Steedle, 2009).

Using CLA as an example, the OLS-based statistical models can be expressed as:

$$\overline{CLA}_{j,F} = \alpha_F + \beta_F(\overline{EAA}_{j,F}) + u_{j,F} \qquad (j = 1,\ldots, n) \qquad (2.1)$$

$$\overline{CLA}_{j,S} = \alpha_S + \beta_S(\overline{EAA}_{j,S}) + u_{j,S} \qquad (j = 1,\ldots, n) \qquad (2.2)$$

where *F* denotes freshman students, *S* denotes senior students, and *j* denotes institution, which is the unit of analysis; $\overline{CLA}_{j,F}$ is the mean CLA test score among freshman students in institution *j*; $\overline{EAA}_{j,F}$ is the mean Entering Academic Ability score[4] (EAA) among freshman students in institution *j*; α_F is the intercept of the freshman regression model; β_F is the slope of the freshman regression model when predicting the CLA mean scores from the EAA mean scores; $u_{j,F}$ is the residual between the observed and expected CLA mean scores from the freshman regression model for institution *j*; and $u_{j,S}$ is the residual between the observed and expected CLA mean scores from the senior regression model for institution *j*.

The institutional value-added score is calculated as, $u_{j,S} - u_{j,F}$, the difference score between the senior and freshman residuals of that institution. Each institution's value-added score is then compared across all institutions that had at least 25 students with both CLA and EAA scores (CAE, 2009). In other words, since it is voluntary for colleges and universities to participate in CLA, a self-selected sample of institutions form the norm group of comparison and the total number of participating institutions is relatively small at the current stage (i.e., around 100-200). Finally, standardized scores (*z*-scores) are calculated among institutional norm groups and used to assign performance levels for each participating institution. Five performance levels are used:

Well Above Expected	2 or more standard errors above the expected score
Above Expected	Between 1 and 2 standard errors above the expected score

[4] It is SAT Math + Verbal, ACT Composite, or Scholastic Level Exam (SLE) score, all placed on the SAT scale.

At Expected	Between -1 and +1 standard errors from the expected score
Below Expected	Between 1 and 2 standard errors below the expected score
Well Below Expected	2 or more standard errors below the expected score

(Source: VSA, 2008a)

Figure 2.2 is a visual presentation of this OLS method (Steedle, 2009). Circles indicate freshmen and squares indicate seniors from example institutions which have sufficient numbers of students with CLA and SAT/ACT scores. Here, each circle or square represents an institution rather than a student, as only average test scores for each institution are used in the OLS method. In general, the average test scores for freshmen and the average test scores for seniors within each institution are not equal, since the circles and squares rarely overlap in the figure. The two lines and the surrounded circles/squares show the expected and the observed relationships between CLA and SAT/ACT scores. ① is the expected average CLA score among freshman students in one institution based on a linear regression when regressing students' average CLA on the average SAT/ACT score across all participating institutions; ② is the expected average CLA score among seniors in that institution; ③ is the expected value added which is the difference in expected CLA scores between the freshmen and seniors tested at that institution based on two fitted regression lines (i.e., ② - ①); ④ is the observed average CLA score among tested freshmen at the institution; ⑤ is the observed average CLA score among tested seniors at the institution; and ⑥ is the difference in observed (unadjusted) CLA scores between the freshmen and seniors tested at that institution (i.e.,

⑤ - ④). The institutional value-added score is calculated as the difference between the observed value added (i.e., ⑥) and the expected value added (i.e., ③). This institutional value-added score is also mathematically equivalent to the difference of two difference scores or residuals of the two fitted regression lines (i.e., (②-⑤)-(①-④) = (②-①)-(⑤-④)). It is then used to determine this institution's value-added performance level. Since two linear regressions based on institutional student samples are used here, it should be noted that the departure from linear regression's assumptions such as linearity and homoscedasticity of the error variances will affect the value-added estimate.

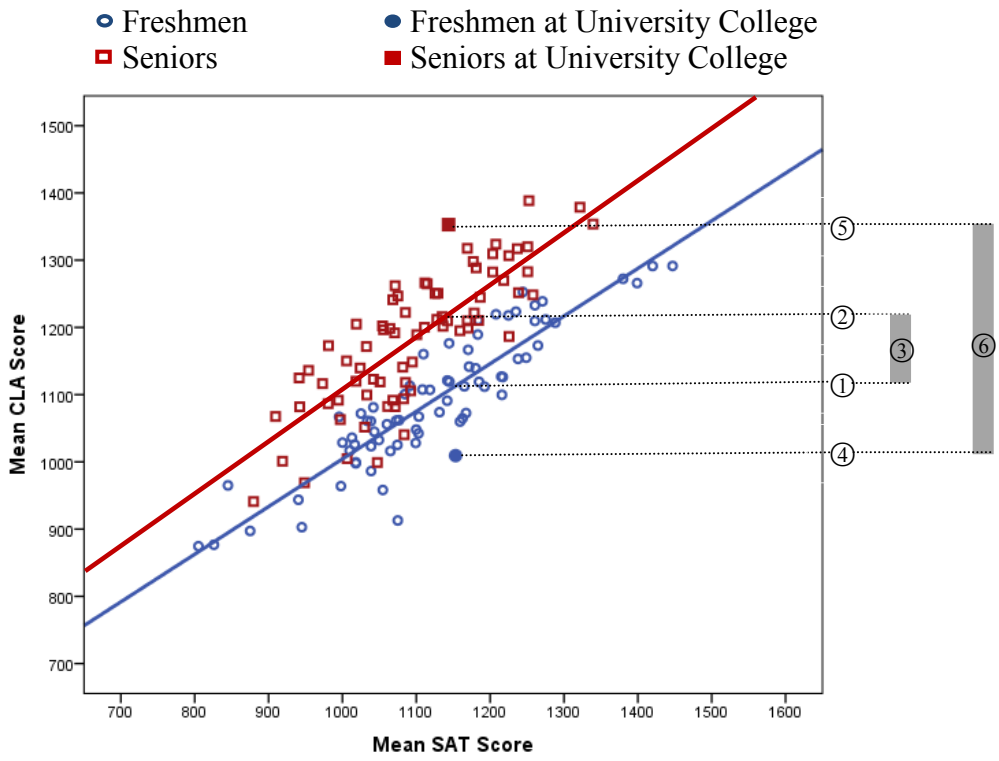


Figure 2.2 The OLS-Based Value-Added Score Estimation Approach[5]

[5] This diagram is similar to one used by Steedle (2009).

The second method adopts a hierarchical linear modeling approach to estimate institutional value-added scores. In a CLA white paper, Steedle (2009) introduced this new method as being able to produce more reliable value-added scores both within and across years, generate school-specific indicators of valued-added score precision, and advance "the quality and interpretability of CLA scores" (p.8). The statistical model of this HLM-based approach is:

Level 1 (Student Level): $CLA_{ij,S} = \beta_{0j} + \beta_{1j}(EAA_{ij,S} - \overline{EAA}_{j,S}) + r_{ij}$

Level 2 (School Level): $\beta_{0j} = \gamma_{00} + \gamma_{01}(\overline{EAA}_{j,S}) + \gamma_{02}(\overline{CLA}_{j,F}) + u_{0j}$

$$\beta_{1j} = \gamma_{10} \tag{2.3}$$

where $CLA_{ij,S}$ is the CLA score of senior student *i* at institution *j*; $EAA_{ij,S}$ is the Entering Academic Ability score of senior student *i* at institution *j*; $\overline{EAA}_{j,S}$ is the mean EAA score among the senior sample at institution *j*; β_{0j} is the mean CLA score for institution *j*, adjusted for the covariate, group mean centered EAA score; β_{1j} is the regression coefficient for institution *j*, associated with the covariate, group mean centered EAA score; r_{ij} is the residual for student *i* in institution *j*, where $r_{ij} \sim N(0, \sigma^2)$ and σ^2 is the variance of the student-level residuals (the pooled within-school variance of CLA scores after controlling for EAA); $\overline{CLA}_{j,F}$ is the mean freshman CLA score at institution *j*; γ_{00} is the intercept for the regression of the adjusted institution mean (equal to grand mean of the CLA scores); γ_{01} is the institution -level value-added equation slope coefficient for senior mean EAA; γ_{02} is the institution -level value-added equation slope coefficient for

freshman mean CLA; γ_{10} is the student-level slope coefficient for EAA (constant across institutions); u_{0j}, the institutional value-added score, is the value-added equation residual for institution *j* (CAE, 2010).

The adoption of HLM was an improvement over the original OLS-based method from several aspects. First, given the hierarchical structure of these student outcome data with students nested within institutions, HLM is a more appropriate method than OLS regression models (Raudenbush & Bryk, 2002). One of the assumptions of OLS models is the independence of observations and their associated residuals (i.e., the difference between the predicted value and the observed value). However, in the case of CLA, students from the same institution are likely to have similar curricular and co-curricular experiences because of the different activities, clubs, and special interest groups available at different postsecondary institutions. Students' scores on CLA are therefore interrelated within each institution due to these unique characteristics. OLS models cannot handle the dependence of observations, but HLM has the advantage of partitioning the between-institution variance from the within-institution variance.

Second, the HLM-based method is able to use student as the unit of analysis. In the OLS-based method, institution is the unit of analysis and students' test scores are aggregated to the institution level (e.g., the circles and squares in Figure 2.2 represent institution average test scores, not individual students' test scores). Student-level information is ignored in the OLS-based method but can be fully utilized in the HLM-based method. Moreover, the results are more reliable given the exceedingly large

number of students compared to the number of institutions in the collected data (Liu, 2011).

Third, the HLM-based method can generate information regarding the precision of estimated school-specific value-added scores (Steedle, 2009). Institutions with large student samples that participated in the tests are estimated more reliably. Such information is also valuable when making comparisons between different institutions' value-added scores.

The last two sections of this chapter discuss these two value-added methodologies of VSA in more detail. They include the advantages and critiques of these value-added estimates, such as reliability and validity issues of the tests, the credibility of using SAT/ACT scores as the only adjusting criterion, small sample sizes within institutions, and student motivation (especially seniors) in finishing a 90-minute written test (Banta, 2006; Banta & Pike, 2007; Hosch, 2010; Pike, 2006; Shermis, 2008). A new longitudinal method with more promising features over the two value-added methods currently used is also introduced.

2.3.4 Student Surveys

Widely used to collect information about institutions and students in higher education, student surveys are a common approach employed to measure noncognitive outcomes such as student engagement and their attitudes and values. In addition to noncognitive related information, students can also be asked about their self-perceived gains in cognitive areas during college, as well as input and process related information

including their demographic characteristics, high school grade point average, and other collegiate experiences.

Many student surveys are administered at the regional and national level. Two of them are the most well-known and largely accepted in the higher education community – the National Survey of Student Engagement (NSSE) and Cooperative Institutional Research Program (CIRP).

NSSE was developed to evaluate the extent to which students are engaged in educationally purposeful activities and what they perceive to be the personal gains from their educational experience. It focuses on both freshman and senior students at four-year colleges and institutions. Since its first official administration in 2000, NSSE has received increased participation and growing attention.[6] Over 600 colleges and universities from the U.S. participated in NSSE 2011.

Reputed to be the oldest and largest study of American higher education, the CIRP survey has been used to inform higher education institutions since 1966. The CIRP survey provides comprehensive information on students' backgrounds and their noncognitive characteristics, such as habits of mind, pluralistic orientation, civic awareness, political and social involvement, and sense of belonging (Sharkness, DeAngelo, & Pryor, 2010). Additionally, the CIRP Freshman Survey, used in conjunction with the Your First College Year Survey (YFCY) and the College Senior Survey (CSS), can provide valuable data for longitudinal studies.

---

[6] Recently, some critiques and concerns have been raised about the design and psychometric properties of the NSSE instruments. For more details, please refer to the special issue of *The Review of Higher Education* in Fall 2011 (Nora, 2011), the response from NSSE (Ewell, McClenney, & McCormick, 2011), and NSSE's commitment to data quality (NSSE, 2011).

Because of this unique advantage of CIRP surveys which supports conducting longitudinal analysis, this study adopts CIRP freshman and senior surveys to explore students' civic engagement levels during college. However, the use of CIRP surveys also brings a limitation to this study. Since the CIRP surveys are designed to comprehensively report students' undergraduate experiences, rather than designed specifically to measure the multidimensional civic engagement construct (e.g., inward attitudes and values towards and outward participations in civic activities), the civic engagement outcome to be investigated in this study does not fully represent this multidimensional characteristic. Based on results presented in later chapters, a scale that measures students' inward attitudes and values toward civic engagement is developed and used in this study.

Other than these two large-scale student surveys, many locally developed and specialized scales also measure students' noncognitive outcomes in higher education. For instance, the Wabash study uses many scales to measure students' noncognitive outcomes, such as Need for Cognition Scale, Miville-Guzman Universality-Diversity Scale, Socially Responsible Leadership Scale, and Ryff Scales of Psychological Well-Being (Center of Inquiry, 2009).

Unlike direct measures of undergraduate students' learning outcomes, student surveys have the advantages of being less costly and relatively convenient for collecting data on a large-scale and on a longitudinal basis. Therefore, many researchers have conducted studies based on survey data. In fact, NSSE and CIRP surveys have been the basis for various college impact studies (Pascarella & Terenzini, 1991, 2005). They have

also been used to understand undergraduate experiences, evaluate student satisfaction, assess students' academic achievement, study student engagement and other specific campus issues, including evaluating programs and guiding teaching practices (e.g., Hurtado, Newman, Tran, & Chang, 2010; Outcalt & Skewes-Cox, 2002; Pascarella, Seifert, & Blaich, 2009; Pike, Kuh, & McCormick, 2011). However, there are concerns regarding student surveys' self-reported nature and the representativeness of the self-selected sample, which are elaborated in the next section.

To summarize, not many instruments have been developed to measure student learning outcomes in higher education. Among the available outcome measures, only a few of them are designed for use regionally or nationwide for accountability purposes. More specifically, accreditation and institution rankings provide basic schemes for making comparisons across different institutions. However, they mainly collect input and process data but lack real learning outcome data on what students know, what they can do, and how much they have gained during college. While standardized assessments directly measure student learning outcomes, most of them are locally developed and administered, so they cannot be used to make cross-institutional comparisons at the national level. Moreover, these instruments only measure cognitive outcomes which are a few of all desired collegial outcomes. Student surveys have the advantage of being a less-costly means of measuring noncognitive outcomes, but also suffer from several concerns.

As summarized in *A Culture of Evidence* (Millett, Stickler, Payne, & Dwyer, 2007), cognitive outcome measurements in higher education have been "well-defined and extensively developed," but that is not the case for noncognitive measures (p.5). At the time the report was published, it only identified one commercially available noncognitive measurement, ACT's WorkKeys assessment in Teamwork. There are other instruments to measure postsecondary outcomes, especially noncognitive characteristics, such as situational judgment tests, interviews, and letters of recommendation. However, they are not widely used and reported at the national level, and, therefore, are outside of the scope of this review.

2.3.5 Challenges in Measuring Higher Education Outcomes

From the above overview of the current status of higher education outcome measures, several challenges emerge involving issues in measurement, instrument, and student samples. Before addressing these challenges, a question that comes to the forefront is "*what* should be measured?" That is, what are the knowledge, skills, and competencies that college graduates should have to survive and thrive in the 21st century? Reviews of the literature at the beginning of this chapter show that higher education outcomes are multidimentional and many different taxonomies have been developed in the past – each of them has had an important influence on the field and has its own proponents. Despite the great efforts that have been made through recent collaborative studies, however, unanimous agreement has not been reached. This issue is even more complex considering the diverse missions and types of higher education institutions.

That is, each institution has its own specific areas of focus among different learning outcome dimensions (see Adelman et al., 2011; Shermis, 2008). Individual institutions can select or develop their own instrument for the purpose of self-evaluation or self-improvement. However, these unique instruments cannot provide evidence in cross-institutional comparisons for the purpose of accountability.

Because higher education outcomes are multidimensional, it is nearly impossible to use a single measurement to make decisions and judgments about postsecondary institutions. This is the first challenge that any effort of measuring higher education outcomes faces today. For example, although higher education institutions can select one of the three standardized assessments in VSA, all three instruments only measure common cognitive outcomes. Similar to the concerns in the K-12 setting, critics worry that postsecondary curriculum will be narrowed if the test results become more high-stakes (Banta, 2007). Moreover, many people disagree that value-added scores, if used for institutional accountability purposes, should be solely based on generic cognitive outcome, such as critical thinking and analytic reasoning, without considering students' discipline-specific knowledge and noncognitive measures that may not be found in the test scores (e.g., Banta & Pike, 2007; Kuh, 2006).

The next challenge is the inconsistent definition of the construct being measured. Concepts and definitions of one construct vary among even a group of experts who extensively work on the same construct. Among those who agree with the importance of critical thinking, different understandings of what this means are highly likely to exist (Pike, 2006). This is also true for noncognitive constructs. Pascarella & Terenzini (2005)

listed the variation in concepts and measurement definitions as the foremost difficult issue "confronting meaningful study of college's effects on students' attitudes, values, and beliefs" (p.271). Another example is the various definitions of the outcome examined in this dissertation study – civic engagement. Many studies have used the term "civic engagement" to mean many different things. As a result, different operationalized items can be used to form the civic outcome, even when the research data come from the same survey instrument.

After the first group of challenges regarding "*what* should be measured," the next group of challenges is "*how* should these outcomes be measured?" These challenges are discussed in the next three sections: instrument related issues, self-selected samples, and student motivation.

Instrument related issues. Although some higher education practitioners are against the idea of standardized cognitive assessments (e.g., Thomson & Douglass, 2009), instrument related concerns are more evident in measuring noncognitive outcomes. As Kyllonen (2005) pointed out, "many policy makers and scientists are skeptical that noncognitive qualities can be measured reliably and in a valid way" (p.2). Because noncognitive measures, especially the items pertaining to personal values and attitudes, have no correct answers or right-answer keys, rather "answers that are correct for each individual" (Messick, 1979, p.283), noncognitive outcomes are normally assessed in the format of self-reported surveys. Students are asked questions such as "How often in the past year did you perform community service?" and "Rate yourself on creativity as compared with the average person your age." The potential problems of such self-reports

include idiosyncratic interpretations student hold for "frequent" and "occasional" participation on campus activities, as well as the different comparison groups each student has in self-rating.

Instrument precision is a related issue as it pertains to detecting the sometimes subtle differences among students. Because all sophisticated statistical models are dependent on reliable and valid tests, researchers have called attention to the measurement accuracy of the assessments (e.g., DePascale, 2006; Gong, Peire, & Dunn, 2006). Most research on cognitive assessments is within the context of NCLB, which includes estimating the reliability and conditional standard errors of measurement of scale scores (Kolen, Hanson & Brennan, 1992), examining the vertical scales and other growth scales used in growth models (Briggs & Weeks, 2009; Tong & Kolen, 2007), the interplay between test difficulty, students' ability, and measurement errors on detecting student growth (Wang, 2009), and standard setting and classification accuracy (Erickan & Julian, 2002; Hambleton & Slater, 1997). However, this issue has not received much attention in the process of designing noncognitive instruments and reporting their results.

As Campbell's Law indicates, increasing reliance on quantitative measures tends to distort the reality that they are trying to measure (Campbell, 1975). Although it is not the current situation, if in the future the survey results are used in high stakes situations to provide school accountability information and make accreditation or resource allocation decisions, unintended consequences such as the "halo effect," faking, or short-term coaching may severely decrease the validity of the survey results. Considerable effort has been put into developing fake-resistant noncognitive assessments (Kyllonen, 2005).

These alternative instruments include reaction time measures, emotional intelligence measures, situational judgment tests, biodata, forced choice, multiple measures, and others' ratings. Some of them are commonly used in certain fields and in local settings. For example, situational judgment tests, which simulate real life scenarios and ask respondents to solve the problem, have been widely accepted in industry (McDaniel, Hartman, Whetzel, & Grubb, 2007). ETS has been using a Standardized Letter of Recommendation to select summer interns and fellowship recipients (Liu, Minsky, Ling, & Kyllonen, 2009). Low-stakes noncognitive assessments (such as the Myers-Briggs Type Indicator® (MBTI®) personality inventory) are also used by institutional career centers to help students select an undergraduate major and provide career counseling (CPP, 2009). However, these innovative noncognitive assessments face other challenges such as being time-consuming and costly and placing heavy burdens on the rater if widely used. Therefore, they are still in the phases of development, validation, or promotion and have not been used in large-scale settings in higher education.

Self-selected Samples. In conducting standardized assessments and surveys on campus, students normally self-select to participate. Taking CLA as an example, research has been carried out to find students' most and least preferable incentives to take the test. Steedle (2010) found that freshmen prefer cash and prizes over other incentives. Yet, the most effective strategy is to make students take the test as a course requirement. In practice, institutions are using various incentives – monetary stipends, gift cards, graduation related gifts, course credit – to encourage students to participate in CLA and meet the recommended sample size of 100 for valid test results. Still, as many

practitioners admit, recruiting students – especially graduating seniors – is by far "the biggest challenge" (Ekman & Pelletier, 2008).

From a statistical point of view, it is highly likely that these voluntary student recruits are not a representative student sample of each individual institution. Without a large representative sample, it is difficult to generalize the study results based on a small group of students to that of the entire student body of the institution.

Student Motivation. Eliciting appropriate student motivation is another challenge these outcome measurements are facing. As student outcome results could be used by federal government and accrediting agencies for accountability purposes, institutions will inevitably feel pressure to present the best student outcomes possible. However, participating students encounter virtually no consequences from the test outcomes. Research has found that students do not take tests or surveys seriously if the results do not have a direct impact on them (O'Neil, Abedi, Miyoshi, & Mastergeorge, 2005; Wise & DeMars, 2005). For example, while the CLA is a two-hour assessment, including a 90-minute essay test and a 30-minute session for a pre-test tutorial and survey, Wall-Smith (2006) have reported that some students simply gave up taking the CLA and left shortly after it started in an urban state university in the Midwest. Using data from a public suburban university in the northern U.S., Hosch (2010) studied the relationship between student performance on the CLA, time spent on the test, and student motivation. Based on descriptive statistics provided in the tables, averaging the CLA 2008 and CLA 2009 two year administrations, freshman students spent 46 minutes and senior students spent 58 minutes on the entire test, which includes the time students spent on the pre-test

tutorial and survey sessions. The researcher also reported that some students quickly left the room; the shortest recorded time was 11 minutes.

The challenges associated with low motivation of participants have been recognized by many researchers (Erwin & Wise, 2002; Wise & DeMars, 2005; Wolf & Smith, 1995). In practice, students are often offered incentives to participate in the test, with the reasons for doing so being two-fold. On the one hand, the institutions are hoping to recruit a large enough student sample. On the other hand, and more importantly, the institutions are hoping the remuneration will act as an internal motivator so that the students take the test more seriously and perform to their abilities. Although monetary incentives have been found to help motivate students to do their best on low-stakes tests and surveys (Braun, Kirsch, & Yamamoto, 2011; Szelényi, Bryant, & Lindholm, 2005), that is not always the case – for example, in the Hosch study mentioned above, senior students were offered a discount or waiver of the entire graduation regalia fee. Providing all students with monetary incentives when they are taking surveys is not realistic and it is not a desired direction. Moreover, many higher education institutions are uneasy doing so.

Steedle (2010) examined the relationship between students' CLA test scores and their performance motivation which was measured by the Student Opinion Survey (Sundre & Moore, 2002). Not surprisingly, student motivation was found to account for 5% of the total variance in CLA score. The incremental R^2 is statistically significant even after controlling for individual students' academic ability (i.e., SAT and ACT scores) upon entry. Hosch (2010) found statistically significant correlations between minutes

spent on the CLA test and CLA scores. Since CLA is a relatively long assessment with only constructed response questions, if students do not try their best to respond, connecting the test results with students' real abilities will be a serious problem. The varying student motivations add difficulty in accurately comparing across institutions. In other words, the desired interpretations of the test scores can be plausibly challenged when students have low or varying motivations when taking the test. Comparing the effort needed to finish a two-hour written test, college students are probably more willingly to finish a survey that lasts less than 30 minutes asking about their opinions and college experiences. Thus, student motivation is more of a concern for the former, which the CLA process uses, rather than the latter, used by surveys like CIRP and NSSE.

2.3.6 Challenges in Estimating Institutional Effectiveness

Using student learning outcomes as an indicator to estimate institutional effectiveness faces many potential challenges. As Pascarella and Terenzini (2005) concluded, "the influence of college or the influence of different collegiate experiences on students is highly dependent on methodological rigor" (p.12). Beyond the challenges in identifying the core constructs to measure, selecting the best instrument and designing scales with acceptable reliability and validity support, and practical challenges in recruiting a representative sample from the target population who also have appropriate motivation, the challenges also lie in the value-added methodology that is adopted to analyze institutional effectiveness. This section starts with an introduction to value-added modeling and proceeds with three value-added methods that can be used to

estimate institutional effectiveness – two methods based on data from a cross-sectional design and one method based on data from a longitudinal design. The advantages and challenges of each method are then discussed.

The main purpose of value-added modeling is to measure the effects of individual teachers, programs, or schools on students' learning outcomes while taking into account students' prior achievement and, sometimes, background characteristics – both at the student and the institution level. In contrast to other types of evaluation models such as status models, cohort-to-cohort change models, and growth models, the advantage of value-added models (VAMs) is that VAMs adopt sophisticated statistical techniques to isolate the approximate contributions of individual schools (or teachers, programs, etc.), so that the value-added scores can be compared across different schools (or teachers, programs, etc.) and be used in the accountability system (Braun, Chudowsky, & Koenig, 2010; Lissitz, Doran, Schafer, & Willhoft, 2006).

Various kinds of VAMs are now used in education. In the K-12 setting, many longitudinal VAMs are used, such as the Educational Value-Added Assessment System (EVAAS), which is marketed by the SAS corporation, and the Dallas Value-Added Accountability System (DVAAS) in the Dallas school system (Sanders, Saxton, & Horn, 1997; SAS, n.d.; Webster & Mendro, 1997). Based on the different assumptions these models make, they differ in the structure of model building. For example, other than controlling for the pre-test measure, many VAMs also control for student background characteristics (such as gender, ethnicity, and socioeconomic status) to take into maximum account possible factors that influence student development (Millman, 1997).

In general, these VAMs can be different in terms of the kinds of statistical adjustments made, the number of years of data employed, and the ways missing data are handled, but they usually require longitudinal data which follow students over time (see Braun, 2005; McCaffrey, Lockwood, Koretz, & Hamilton, 2003; and Wainer, 2004, Journal of Educational and Behavioral Statistics special issue). Therefore, strictly speaking, cross-sectional data which only include students' learning outcomes at one time point is not sufficient for building VAMs (some are contextualized attainment models; see OECD, 2008a). In a traditional cross-sectional study, data are collected from one group of the sample at one time point. In the context of measuring student learning outcomes in higher education, students are examined at a single time point, which can be either before entering college, at the end of the first college year, or during the last year of college. Information collected can include students' academic knowledge, skills, their attitudes toward certain issues, co-curricular activities they participated, and their expectations and satisfactions toward the college. The analysis normally involves correlation and single regression analysis to examine the relationships among variables at that time point. Although such a cross-sectional design is easy to conduct and cost-effective, the collected information cannot say much about whether students are changing, how they are changing, or explain the difference between two institutions' educational effectiveness (Astin, 2003).

For the purpose of estimating institutional effectiveness, the three methods discussed next are slightly different from the traditional cross-sectional method. Data for the first two methods are still collected through a cross-sectional design, but efforts are

made to consider student outcomes of both freshman and senior students. The first method (hereinafter, Method 1) is the original VSA method now used by the ETS Proficiency Profile and ACT's CAAP. In the data collection procedure, both freshman and senior students attending the same institution are assessed at a single time point, and the final data include information collected from multiple institutions. During data analysis, students' scores are aggregated to the institution level. The institution is then treated as the unit of analysis to calculate the residuals in predicting the achievement outcome. After obtaining the residuals separately for freshman and senior students, the difference between the two residual scores is the estimated value-added score (ACT, 2009; Liu, 2008).

The second method (hereinafter, Method 2) is the new method CLA developed and uses in its value-added calculation. Its statistical model is similar to a value-added model using true longitudinal data. Specifically, a two-level HLM is used to predict senior students' academic ability scores while adjusting for the mean of the test scores among freshmen for each institution at the institution level (see equation 2.3). Because this method attempts to control for students' prior achievement – although with a proxy – it is different from a pure cross-sectional method. However, because the freshman and senior students are different groups of students who study in the institution at the same time, the adjusted school average score is only a proxy of, but not real prior achievement of, the group of senior students.

The third method (hereinafter, Method 3) uses true longitudinal data which follows the same individuals at two time points. Menard (1991) has broadly categorized

four types of longitudinal designs, namely, total population design, repeated cross-section design, revolving panel design, and longitudinal panel design. Using longitudinal CIRP data that track undergraduate students over time, the longitudinal method in this study only refers to the longitudinal panel design in Menard's definition. It is the most familiar type of longitudinal design when the same group of individuals is measured in each time period, although attrition is likely. Specifically, it is different from a cohort analysis which includes multiple cross-sectional studies (e.g., Lawlor, Clark, Ronalds, & Leon, 2006; Marcon, 1999).

Each of the above methods has advantages and disadvantages. In general, cross-sectional designs are easy to conduct, less costly, can produce results more quickly, and can usually collect information from a large sample that does not suffer from respondents' attrition. However, it has a snapshot characteristic and the status data collected limit its use in modeling students learning and changes over time.

In the larger education research field, the argument between cross-sectional and longitudinal methodologies seems to have been settled in favor of a longitudinal method. In the K-12 setting, to solve the problems derived from the snapshot characteristics of Adequate Yearly Progress in the NCLB accountability system, the U.S. Department of Education has made several attempts to improve the system by using longitudinal methods. Some examples are the approval of the Growth Model Pilot Program in 2006, and Delaware and Tennessee becoming first round winners in the Race To The Top competition – both Delaware and Tennessee have demonstrated a strong capacity for

adopting longitudinal data to link student achievement to the teacher and school accountability system.

In the field of higher education, researchers have long recognized the importance of longitudinal designs from college impact studies (Astin, 1970, 1993a; Astin & Lee, 2003; Pascarella, 2006; Seifert, Pascarella, Erkel, & Goodman, 2010). Using an empirical example, Seifert, Pascarella, et al. (2010) demonstrated that cross-sectional designs adopt inaccurate proxies that are different from a true baseline pretest; thus, they provide estimates that are confounded with selection bias. On the other hand, longitudinal designs "maximiz[e] the internal validity" and provide the most accurate estimate of college impact studies when it is unlikely to design randomized experiments in practice (p.5). The authors further concluded that, in the absence of experimental designs, "there is no substitute for the 'gold standard' that longitudinal pretest-posttest studies furnish in accurately assessing how students learn and change" (p.14). Pascarella (2006) also suggested that carrying out multiple small-scale longitudinal studies may be more valuable than analyzing data from several institutions using cross-sectional designs. In Astin's (1970, 1993a) I-E-O model, he emphasized incorporating students' pre-measures in the study to minimize input bias and the importance of avoiding one-shot cross-sectional assessments. The Cooperative Institutional Research Program actually originated from this notion after recognizing the importance of longitudinal designs (Astin, 2003).

In practice, the two value-added estimation methods used by VSA inherit the disadvantages of a cross-sectional design. Because the learning outcomes are from

different groups of freshman and senior students, the precision of its measurement of the student outcome and the estimated institutional value-added scores is a concern. The comparability between the two different groups of students could become an issue. Specifically, different characteristics of the two student groups such as different student drop-out and transfer rates, differences in the student outcome of interest, and the various differences they bring with them before matriculation all add confounding factors in the value-added results.

Moreover, the institution is the unit of analysis in Method 1, which means all the student-level information is ignored. This also creates a potential for an "ecological fallacy" (Robinson,
1950) when the conclusions drawn at the institutional-level are incorrectly applied to all individual students in that institution. Since students greatly outnumber institutions, more reliable results could be obtained had the student-level information been used in the value-added analysis (Liu, 2011). As Method 2 is a new method, little research has been done on it. Thus, there is a lack of empirical evidence comparing its results and those from longitudinal designs.

These two methods have potential flaws that call for further examination. However, longitudinal studies are not challenge-free. Many existing longitudinal studies, such as the Wabash study, face problems from student attrition and lower than desirable sample sizes (Klein, 2009; Seifert, Goodman, King, & Baxter Magolda, 2010). In addition to the attrition of participants, another practical concern is that longitudinal studies are time-consuming. To measure institutional value-added, a minimum of four

years are required. This also means a delay in obtaining needed information and revealing study results. Other sources of error, such as the possibility of selective sampling and test-retest effects can also threaten the validity of the study (Bauer, 2004).

Moreover, the ultimate challenge in the use of any VAM lies in the inferences one can make from those value-added scores, as researchers hope to link student outcomes with institutional quality (or institutional effectiveness) in this era of accountability.[7] As Braun (2005) noted, due to the lack of randomization, causal attributions cannot be confidently made about the quality of teaching (or education program or school) – no matter how complex the statistical model is and how sophisticated the method of analysis is. The challenge is that many attributes not measured may be associated with these low stakes measures. Therefore, caution is needed before building a link between student outcomes and institutional effectiveness or making conclusions about institutions for accountability purposes. Shavelson (2010) described accountability as "a delicate instrument." The use of low-stakes measures in high-stakes situations can bring unintended negative consequences which are contrary to the original positive purposes of test designers. The results from student learning outcomes ought to be used properly "to inform decision making and improve teaching and learning," rather than with high-stakes accountability (Kuh & Ikenberry, 2009, p.4).

In general, the main purpose of this study is to explore some of the above mentioned challenges in estimating institutional effectiveness. The two areas of concentration include focusing on civic engagement, a noncognitive outcome, and

[7] For more challenges on VAMs, see Braun et al. (2010).

examining the differences in results between cross-sectional and longitudinal value-added methodologies. In the next chapter, the methodology to be used in this dissertation is presented.

CHAPTER 3. RESEARCH DESIGN

3.1 Data Description

3.1.1 Instruments

Data in this study comprise two major components: (1) the CIRP undergraduate student survey data from HERI at UCLA; and (2) institutions' reported data from NCES' IPEDS. Specifically, two survey instruments conducted by HERI are used – The Freshman Survey (TFS), and College Senior Survey (CSS). These survey instruments are widely used and are crucial in understanding higher education. They are introduced first.

The CIRP is the "largest and longest-running longitudinal study" of American higher education (Astin, 2003, p.28). Since its first use in 1966, the CIRP surveys have collected data from over 15 million students and over 1,900 institutions (Pryor, Hurtado, DeAngelo, Blake, & Tran, 2010). ***The TFS*** is generally administered at the start of a student's freshman year in college and collects data about student background characteristics (e.g., gender, race/ethnicity, parental income, religion), students' experiences in high school, their expectations of college, self-ratings in comparison with others of the same age (e.g., academic ability, self-confidence, spirituality), attitudes, and values. The TFS data also include some institutional-level information such as institutional type and selectivity.

The CSS is generally administered to senior students before graduating from college. It "connects academic, civic, and diversity outcomes with a comprehensive set

of college experiences to measure the impact of college" (HERI, 2011). While the CSS repeats most of the questions in the TFS regarding students' attitudes, values, and self-ratings, it also collects information about students' college experiences, their satisfaction towards the college, as well as self-descriptions in comparison with when they first started college. Because of the large number of common items in the TFS and CSS, the CIRP surveys offer a unique advantage of being able to carry out a longitudinal study and have been used to track student development during college in many studies (see Astin, 1993a; Pascarella & Terenzini, 2005). In addition, the CSS has been selected and incorporated into the VSA College Portrait[8] as one of the four survey instruments[9] to provide comparable information about student experiences and perceptions, but it has not been used as part of the student outcome measures to calculate college value-added scores (VSA, 2008b).

IPEDS is the Integrated Postsecondary Education Data System. Conducted by the U.S. Department of Education's National Center for Education Statistics, IPEDS is a system of interrelated surveys collecting information from "every college, university, and technical and vocational institution that participates in the federal student financial aid programs" (NCES, 2011). IPEDS data include information such as institutional characteristics, resources, prices, student enrollment, and financial aid. In this study, several variables are selected to provide information about institutional characteristics

[8] The College Portrait is "a source of basic, comparable information about public colleges and institutions presented in a common, user-friendly format… College Portrait is the name of the 5-page report designed to present the VSA data to the public" (College Portrait, 2009).
[9] The other three survey instruments are: the College Student Experiences Questionnaire, the National Survey of Student Engagement, and University of California Undergraduate Experience Survey (VSA, 2010).

(e.g., location, size and setting, Carnegie Classifications), student enrollment (e.g., full-time equivalent students' gender and race/ethnicity), and faculty demographics (e.g., full-time equivalent faculty members' gender and race/ethnicity). These institution-related variables from IPEDS are merged with the CIRP survey data for further analysis.

3.1.2 Data Source and Sample

To investigate the differences between cross-sectional and longitudinal study designs, this study draws data from four survey administrations: TFS and CSS in both 2002 and 2006 (These survey instruments are provided in Appendix A). Students who participated in both TFS in 2002 and CSS in 2006 are linked through the same ID number in the fifth dataset. An anonymous school ID is also included in all individual datasets, which helps identify the schools that participated in multiple survey administrations. To maintain school confidentiality, IPEDS data were pre-merged with CIRP surveys by HERI staff.

Since the handling of missing data is a critical issue in most large-scale studies, the techniques of dealing with missing data in this study are introduced here. Because missing data can be problematic in making inferences from the study results, the amount of missing data was examined first. Among the variables of research interests, most variables, such as gender, religious preference, and parental education levels, have less than 3% missing. However, there are also three variables missing between 10-20% of values in different survey administrations. They are: students' educational aspirations, SAT/ACT scores, and family income (see Table 3.1). Due to the relatively large amount

missing in some of the variables of interest, multiple imputation was considered in the study.

Multiple imputation is one of the modern methods for handling incomplete data (Little & Rubin, 2002; Schafer, 1997). A major advantage of it over simple imputation methods is that multiple imputation introduces uncertainty in parameter estimations by creating multiple sets of complete data in which more than one value for each missing case is provided. Moreover, because missing cases are imputed using information from other available variables in the data, it is assumed that data are missing at random (MAR). That is, the missingness of a variable does not depend on the value of that variable after controlling for other observed variables. This is a much weaker assumption than missing completely at random (MCAR) and more realistic for real-world situations.

In the data screening process, when an incomplete variable was found, the associations between the missingness of the variable and a group of variables in the datasets were investigated. MAR assumption can be assumed for all incomplete variables of research interest in the study. For example, in the longitudinal dataset, 12% of the students have missing information on their SAT or ACT test scores (see Table 3.1). Because not all postsecondary institutions require SAT/ACT score, it is possible that some students did not take the test. It is also possible that some students may not be willing to report their test scores. In the preliminary analysis, the variable indicating whether a student had an incomplete SAT/ACT score was found to be associated with some variables, but not others, in the dataset. Compared to their peers, the students with missing SAT/ACT information had lower high school and college GPAs and their parents

had lower education levels. However, the missingness on SAT/ACT is not associated with students' civic engagement measure in both freshman and senior years. These detected associations support the assumption that SAT/ACT scores are MAR after controlling for these observed variables.

Therefore, this study used the multiple imputation method to estimate incomplete cases for most variables, with the only exception being the civic engagement measures (see further discussion in section 3.2.1, stage three). To maximize the information available in the datasets, variables used in the multiple imputations were civic engagement measures created through a three-stage analysis (see section 3.2.1 for more details), student demographics and family background characteristics, high school/college academic outcomes and experiences, and all other variables associated with the civic engagement outcome indicator. All imputation models also included school level variables to maintain the original associations and to take into account the hierarchical structure of the data (Little & Rubin, 2002). Based on Rubin (1987), when the fraction of missing information is around 10-20%, three to five imputations are sufficient to obtain relative efficiency (97-99%) of the estimates (see Rubin, 1987, p.114). In the end, a total of five multiply imputed datasets for each CIRP data set were created using Stata 11 through the Markov Chain Monte Carlo method. The percentages of cases imputed in each of the five datasets are summarized in Table 3.1.

Table 3.1 Percentage of Cases Imputed in the Five Datasets

	TFS02-CSS06 (%)	TFS02 (%)	TFS06 (%)	CSS02 (%)	CSS06 (%)
Gender	-				
Race/Ethnicity	-	-	-	-	-
Religious preference	0.1				
Native language	-				
Parental income	10.9				
Father education	1.1				
Mother education	0.5				
SAT/ACT	12.0				
High school GPA	1.2				
College GPA	0.5				
Educational aspiration	2.6	17.4	12.0	18.5	17.6
Performed volunteer work	0.6	0.7	0.9	0.6	0.6
Voted in a student election	0.8	1.2	1.3	2	0.9
Student clubs/groups	1.1	1.7	2.8	0.9	0.9

Note. "-" means the variable had no missing cases. Variables in the empty cells were not used in the analysis, so they were not imputed.

It should be noted that multiply imputed data have not been incorporated with all available statistical techniques in subsequent data analysis. Factor analysis is one of them. Truxillo (2005) has suggested a maximum likelihood approach with the expectation-maximization (EM) algorithm to estimate the covariance matrix. She compared eight different missing data handling methods and found that the covariance estimates from the EM algorithm produced superior results. Applying this method[10], the derived variable

[10] A more detailed Stata command on applying this method is available at http://www.ats.ucla.edu/stat/stata/faq/factor_missing.htm

representing the socioeconomic status was a factor score constructed from the covariance matrix of three CIRP items: parental income, father's education level, and mother's education level.

Due to the nesting structure of the datasets, hierarchical linear models were used in the estimations to answer the research questions with HLM 6.08 (Raudenbush, Bryk, & Congdon, 2004). One thing to notice is that HLM requires complete data at the second level. This means that missing cases at the institution level need to be removed or replaced. Fortunately, for most institutional variables, the number of missing values is negligible (less than 1%). Of the missing values, the primary cause is the failure of merging CIRP and IPEDS data. For example, six out of 697 (0.9%) institutions in TFS02 and four out of 614 institutions (0.7%) in TFS06 were unable to be matched with data from IPEDS. These institutions were not included in the analyses that use variables from IPEDS. Also, some schools do not require SAT or ACT scores in admissions, which results in about 10% missing on the institutional selectivity variable. These institutions were also removed when the institutional selectivity variable was involved in the analysis. Thus, the number of students and institutions are subject to change in different analyses.

In the data preparation process, some students were found to have transferred schools between their freshman and senior years; some responded in the 2006 senior survey that they entered their current college before or after 2002. Therefore, the following criteria were also used to select the students from the longitudinal dataset:

- The student's record in TFS02 is matched with a record in CSS06;
- The student was in the same institutions in 2002 and 2006;

- The student did not transfer from a different 4-year college or a community college to the current one;
- The student responded in CSS06 that s/he first entered the current college in 2002;
- The student was a first-time, full-time student in 2002 and a full-time undergraduate student (not part-time student or graduate student) in 2006
- The highest degree the student would have earned as of June 2006 does not include Doctoral and professional degrees (such as Ph.D., Ed.D., M.D., and J.D.)

Applying these criteria, about 30% of the students were removed from the dataset.

Finally, an institution needs to meet one additional criterion before being included in the analyses. That is, the institution should have at least 20 students who participated in the survey(s). Because a representative student sample is necessary to make valid inferences about the participating institution as a whole, test developers in the VSA, for example, set basic, minimum standards for student sample sizes (VSA, 2008a). CAE sets it as 100 each for freshmen and seniors.[11] ETS and ACT set it as a minimum of 200 each for freshmen and seniors (Shulenburger & Keller, 2009). To avoid losing available institutions, the minimum sample size first is set as 20 each for freshmen and seniors in this study. Next, to take account of differences in the number of respondents in different

[11] This is only a recommended sample size. In CAE's final report for the 2009-2010 assessment cycle, the analyses included results from "all institutions, regardless of sample size and sampling strategy" (CAE, 2010, p.5). In their 2008-2009 final report, individual institution's percentile ranks for freshmen/seniors were calculated among institutions "where at least 25 students had both a CLA and EAA score in fall 2008 and/or spring 2009" (CAE, 2009, p.3).

institutions, the value-added analyses are replicated for the set of institutions that have a minimum of 100 each for freshmen and seniors.

After adopting the above selection criteria with a minimum of 20 students surveyed in each institution, the total number of institutions and their full-time, first-time undergraduate students available in the data sets are summarized in Table 3.2. It shows that, in both 2002 and 2006, a large number of institutions participated in the freshman surveys. But less than a quarter of the institutions participated in the senior surveys. Moreover, the average number of students within institutions that participated in the freshman surveys was over twice the size of the average number that participated in the senior surveys.

Table 3.2 Number of Institutions and Students in the Data Sets

Survey	Institutions	Students[a]	Avg. number of students/institution
TFS2002	695	379,469	546
TFS2006	613	365,991	597
CSS2002	170	35,179	207
CSS2006	139	28,557	205

[a] TFS data only include full-time, first-time freshmen; CSS data only include full-time undergraduate students.

Table 3.3 is a comparison among the distributions of these institutions' basic Carnegie classification and the distribution of the overall number of institutions in the Carnegie classification. It shows that a majority of the CIRP participants are four-year institutions, including baccalaureate, master's, and research universities and colleges, which is an over-representation of these types of institutions in the Carnegie

Classification. Moreover, associate's colleges and special focus institutions are under-represented in the CIRP surveys. This matches well with the purpose of this study, which focuses on freshman-to-senior student development in four-year institutions.

Table 3.3 Distributions of the Participating Institutions and the Overall Distributions based on Carnegie Classification 2005: Basic

Institution Type	TFS2002 (%)	TFS2006 (%)	CSS2002 (%)	CSS2006 (%)	Overall[a] (%)
Associate's Colleges	21 (3.0)	13 (2.1)	2 (1.2)	-	1740 (41.0)
Baccalaureate Colleges	282 (40.6)	256 (41.8)	82 (48.2)	66 (47.5)	755 (17.8)
Research Universities	118 (17.0)	102 (16.6)	18 (10.6)	13 (9.4)	281 (6.6)
Master's Colleges and Universities	246 (35.4)	214 (34.9)	64 (37.6)	59 (42.4)	656 (15.5)
Special Focus Institutions[b]	23 (3.3)	23 (3.8)	3 (1.8)	1 (0.7)	792 (18.7)
Others[c]	5 (0.7)	5 (0.8)	1 (0.6)	-	18 (0.4)
Total	695	613	170	139	4242

[a] The institutions not in the Carnegie universe (not accredited or nondegree-granting) are excluded. Source: IPEDS Data Center, Institutional Characteristics Survey – Directory Information, 2008.
[b] Special focus institutions include theological seminaries, medical schools and medical centers, schools of engineering, schools of law, etc.
[c] Others include institutions that are not classified and not known.

With the purpose of estimating institutional effectiveness and comparing different estimating methods, most analyses in this study examined data from multiple CIRP surveys administered to freshman and senior students – from either the same or different student cohorts. Table 3.4 lists the number of institutions that participated in more than one survey administrations.

Using the three analyzing methods first introduced in section 2.3.6 (please also refer to section 3.3.4 for more detailed information about Methods 1-3), different numbers of institutions were included in different analyses. For example, the 90 institutions that participated in the freshman survey in 2002 and the senior survey in 2006 were examined in Method 3, the longitudinal value-added method. In Method 1 and Method 2, the cross-sectional methods, institutional value-added scores in 2002 and in 2006 were calculated and compared among 86 common institutions. When examining the similarities and differences across all three methods, 66 institutions' value-added estimates were investigated.

Table 3.4 Number of Common Institutions

	Surveys	Institutions
1	TFS2002 & CSS2006	90
2	TFS2002 & CSS2002	152
3	TFS2006 & CSS2006	120
4	Surveys in 2 & 3	86
5	Surveys in 1, 2, & 3	66

3.2 Variables

3.2.1 Outcome Variable

The outcome variable for all three research questions of the study is an indicator of the construct of interest – civic engagement. Incorporating the methods of reporting background questionnaire data used in the International Civic and Citizenship Education

Study (ICCS), the Program for International Student Assessment (PISA), and the Trends in International Mathematics and Science Study (TIMSS), this important variable was derived through three stages with the purpose of minimizing measurement error (Martin & Preuschoff, 2008; OECD, 2008b; Schulz & Brese, 2008). The three stages are presented after a brief overview of current measures of civic engagement.

In previous studies examining civic engagement, multiple dimensions have been identified and various indicators have been used to measure civic engagement. For example, the ICCS is a multi-year, international comparative study by the International Association for the Evaluation of Educational Achievement (IEA). The most recent 2009 ICCS measured the content dimension, affective-behavioral dimension, and cognitive dimension of "civic knowledge, attitudes, and actions of 14-year-olds and upper secondary school students" through a cognitive test, a set of student, teacher, and school questionnaires, and a national contexts survey (Schulz, Ainley, Fraillon, Kerr, & Losito, 2010, p.13). The international student questionnaire was designed specifically to measure several key indicators of civic engagement, including "students' interest in political and social issues, civic participation in the wider community and at school, expected participation in national elections, and expected participation in political activities" (Schulz et al., 2010, p.45).

In the literature examining civic engagement related constructs using CIRP data, a group of survey items have been identified using exploratory factor analysis. For example, Pascarella et al. (1988) derived a factor score from six items to measure humanitarian/civic involvement values. These items are: the importance of (1) becoming

involved in programs to clean up the environment, (2) helping others who are in difficulty, (3) participating in a community action program, (4) becoming a community leader, (5) influencing social values, and (6) influencing the political structure. Building upon this work, Rhee and Dey (1996) found a similar 7-item scale to measure students' civic value. Instead of using the item "becoming a community leader," two items were added – promoting racial understanding and developing a meaningful philosophy of life. Many other research studies also used similar items (e.g., Astin & Antonio, 2004; Astin, Vogelgesang, Ikeda, & Yee, 2000; Sax & Astin, 1998). Next, the three-stage analyzing method in constructing the civic outcome is presented.

Stage One: Exploratory Factor Analysis and Unidimensionality

The civic engagement construct is multidimensional; however, the above items only reflect the attitudinal aspect of the construct. Therefore, a larger group of items from the CIRP surveys related to students' attitudes toward, and participation in, civic activities were examined in the first stage of the analysis. First, all items related to the civic construct and measured in all four CIRP survey administrations (i.e., TFS2002, TFS2006, CSS2002, and CSS2006) were selected. Then, they were analyzed using exploratory factor analysis. The goal of the analyses in this stage is to build a scale that reliably and consistently measures civic engagement across different student samples. So the same group of items is needed to create the civic engagement scale. All items should also have relatively high loadings on the extracted factor. Therefore, in these exploratory analyses, if an item was found to have a factor loading less than 0.3, this item was

removed and the rest of the items were re-estimated. This iterative process was carried out in each of the four survey administrations. Next, Cronbach's alpha reliability coefficient of the extracted factor was reported. It is hoped that there is a high internal consistency among the group of items so that the civic engagement measure in the study is reliable. Because multiple factors were identified among the selected items (see section 4.1.1 for more detailed results), further exploratory factor analyses were carried out on each extracted factor to confirm its unidimensionality. Decisions were then made on item selection for generating the civic engagement scale based on the results.

Stage Two: Confirmatory Factor Analysis and Multisample Confirmatory Factor Analysis

On the basis of substantive theory and prior research on civic engagement, confirmatory factor analysis (CFA) was conducted with LISREL 8.80 (Jöreskog & Sörbom, 2006) to test the nature of the civic engagement construct. In view of the categorical survey items, weighted least squares estimation was used and polychoric correlation and asymptotic covariance matrices were calculated in the process (Jöreskog, 2005). Moreover, multisample CFA was used to determine whether CIRP surveys measure the same constructs across different levels of undergraduate students and different administrations. As suggested by Jöreskog and Sörbom (2001), multisample analyses can include a series of steps, ranging from testing the equivalence of the covariance matrices of the observed items to factor pattern, factor loadings, measurement errors, and the variances and covariances of the latent constructs or factors. This study

examined two of the most fundamental steps, factor pattern invariance and factor loading invariance, to provide validity support to the civic construct used in this study.

Goodness-of-fit indices measure the extent to which the *a priori* structure fits the data. In the CFA models, model fit was assessed through the root mean square error of approximation (RMSEA), the standardized root mean square residual (SRMR), and the Tucker Lewis Index (TLI). In the multisample CFA models, χ^2 and $\Delta\chi^2$, along with the degrees of freedom and statistical significance, were reported to compare competing nested models. The rule of thumb, supported by simulation research, suggests that RMSEAs below 0.06, SRMRs below 0.08, and TLIs greater than 0.95 indicate a good fit (Hu & Bentler, 1999). These indices were used as measures of goodness-of-fit in the analyses. Because fit indices, such as χ^2 and $\Delta\chi^2$, are sample size dependent, which often result in rejecting good models with large sample sizes, the individual sample CFA models were estimated from 500 randomly selected cases of the data and the entire sample's correlation and covariance matrices. The multisample analysis, therefore, included 2,000 cases in total.

Stage Three: Item Response Theory (IRT) Scaling Analysis

After the first two stages, the next step is to derive a scale score for selected items that measures college students' civic engagement. The 1-Parameter IRT (Rasch) Partial Credit Model (Masters & Wright, 1997) has been successfully used and proposed for use in reporting background questionnaire data in many large-scale international comparative studies (OECD, 2008b; Preuschoff, 2010; Schulz et al., 2010). Using IRT scaling,

researchers have the advantage of examining the fit statistics and the scaling properties of categorical items. Rasch measurement software (such as Winsteps®) also handles missing responses well (Lincare, 2011). Therefore, the civic engagement scale measured in CIRP surveys was constructed through the Winsteps Rasch measurement software (Lincare, 2006). The Rasch partial credit model takes the form of

$$P_{nix}(\theta) = \frac{\exp \sum_{j=0}^{x} (\theta_n - \delta_i + \tau_{ij})}{\sum_{h=0}^{m_i} \exp \sum_{j=0}^{h} (\theta_n - \delta_i + \tau_{ij})}, \; x_i = 0, 1, \ldots, m_i \qquad (3.1)$$

where $P_{nix}(\theta)$ is the probability of person *n* scoring *x* on item *i*. θ_n is the person's latent trait. The item parameter δ_i denotes the location of the items on the latent trait continuum and is often referred to as the item difficulty parameter. τ_{ij} (j = 0, 1, …, m) is a step parameter indicating the difficulty of reaching step *j* for item *i*.

It should be noted that the missing responses from the eight items on the civic engagement scale were treated differently from the rest of the variables used in this study. Although Winsteps can handle missing responses by examining the response pattern of the subject on other items (Lincare, 2011), a small percentage of students were removed from the scaling analysis when they had missing responses to all eight items on the civic engagement scale (e.g., 1.5% in CSS02 and 1.4% in CSS06) or had missing responses on more than two of the items (e.g., 0.4% in CSS02 and 0.3% in CSS06). As imputation is not required in Rasch estimations (Lincare, 2011), multiple imputation technique was not used to impute civic engagement related items.

The Rasch analysis was carried out through two steps. After the examination of the construct unidimensionality in the exploratory factor analysis, item calibration first proceeded in order to obtain item parameters. Considering the hierarchical structure of the data with students nested in institutions, and the differences in the number of student respondents in each institution, a stratified random sample was used in the item joint calibration process. Specifically, a sub-sample of 2,000 students was randomly selected from each survey administration. In the resulting random sample, the number of students within each institution is proportional to the total sample size of that institution (i.e., the number of students who responded to the CIRP survey in that institution). A final sample of 8,000 students from all four survey administrations was used simultaneously to estimate item parameters in the Rasch partial credit model. Using pre-calibrated (i.e., fixed) item parameters, the next step derived IRT scale scores through the maximum likelihood estimation (MLE) method[12] for all students who participated in one of the four CIRP survey administrations.

3.2.2 Student-level Predictors

At the student level, four types of variables were considered. They are student and family background, high school educational outcomes and educational aspirations, college educational outcome and experiences, and a student profile variable. Table 3.5

[12] In contrast with ICCS and PISA which use Weighted Likelihood Estimation (WLE), MLE is used in this study. Because Winsteps believes that "transposing the data matrix (exchanging the persons and items) should impact estimation as little as possible" and "estimating the person and item measures from a free analysis should produce the same results." In practice, the difference between these two methods is small and "implications of this for practical applications are not clear." Source: Rasch Measurement Transactions. (2007). *Warm (Maximum) Likelihood Estimates, 21*(1).

presents and describes these student-level variables in detail. As the table shows, student and family background variables include student demographic characteristics (such as gender, race/ethnicity, and native language) and family background characteristics (such as an indicator of socioeconomic status). Educational outcomes include students' grade point average in high school and college, and their college entrance test scores (i.e., SAT and ACT[13]). College experiences focus on activities students took part in during college.

The literature has consistently found an association between SAT/ACT scores and students' cognitive outcomes in college (Kobrin, Patterson, Shaw, Mattern, & Barbuti, 2008; Noble & Sawyer, 2002). However, the association between noncognitive variables and students' academic achievement is found to be weak because of the relative independence between these two types of variables (Knox et al., 1993; Robbins et al., 2006). Therefore, with the aim of capturing college students' noncognitive traits, a student profile variable was constructed.

[13] ACT scores were converted to SAT scores using SAT-ACT Concordance Tables (College Board, 2008).

Table 3.5 Descriptions of Student-Level Independent Variables

Variables	Descriptions
Civic engagement in the freshman year	Continuous variable with a mean around 50 and standard deviation around 10 (these statistics can vary in different datasets)
Student Demographics & Family Background	
Gender	Dichotomous variable: 1=Male, 0=Female
Race/Ethnicity	A group of four dummy variables: White/Caucasian (reference group), African American, Asian American, Hispanic American, and Others
Native Language	Dichotomous variable: 1=English, 0=Other languages
Religious Preference	A group of three dummy variables: Roman Catholic (reference group), Other Christian, Other religion, No religion
Socioeconomic Status	Factor score of: Father's Education (8-category variable from "Grammar school or less" to "Graduate degree"); Mother's Education (8-category variable from "Grammar school or less" to "Graduate degree"); and Parental Income (13-category variable ranging from "Less than $10,000" to "$200,000 or more")
High School Educational Outcomes & Educational Aspirations	
SAT/ACT	Continuous variable: SAT Verbal+Math score, or ACT composite score (converted on the SAT score scale)
High school GPA	Continuous variable on a 1-4 scale; number grades were transferred from letter grades: A or A+ = 4.00, A- = 3.67, B+ = 3.33, B = 3.00, B- = 2.67, C+ = 2.33, C = 2.00, and D = 1.00
Educational Aspiration	A group of six dummy variables: Bachelor's Degree (reference group), Master's Degree, Professional Degrees in Law (J.D.), Medicine (M.D., D.O., D.D.S., D.V.M.), and Divinity (M.Div.), Doctoral degree, and Others
College Educational Outcomes and Experiences	
College GPA	Continuous variable on a 1-4 scale; number grades were transferred from letter grades: A = 4.0, A- or B+ = 3.5, B = 3.0, B- or C+ = 2.5, C = 2.0, C- or less = 1
College Experiences	Sum of three items related to civic participations during college (base on the analysis in section 4.1.1); a composite variable ranges from 3 to 9
Student Profile Variable	
Student Profile	A group of four dummy variables: Status striver, Artists, Hedonists Religious-oriented, and Disengaged (reference group) Created on the basis of 34 items, such as Importance of becoming accomplished in one of the performing arts (acting, dancing, etc.)

Kuh, Hu, and Vesper (2000) reviewed a number of student profiles developed since the 1960's to organize college students into peer groups. Although researchers used different inquiry approaches, they found that the independently developed student profiles were reasonably stable over time. For example, in Clark and Trow's (1966) most frequently cited work, four major groups were identified: (1) Academic; (2) Collegiate; (3) Vocational; and (4) Non-Conformist. Using multi-institutional CIRP data, Astin (1993b) derived seven similar types of students through factor analysis: (1) Scholar, (2) Leader, (3) Hedonist, (4) Status Striver, (5) Social Activist, (6) Artist, and (7) Uncommitted. These student profile studies, especially the more recent ones, predominantly use factor analysis and/or cluster analysis (e.g., Luo & Jamieson-Drake, 2005).

Using CIRP items regarding students' activities, self-ratings, and perceived goals in life in the freshman survey, a student profile variable was created in the TFS02-CSS06 longitudinal dataset and used as a covariate when predicting senior students' civic engagement level (more details are in Section 3.3 Analytic Strategy). One distinction of this study is the use of latent class analysis (LCA). Unlike better-known factor analysis and cluster analysis, LCA only drew attention after the development of advanced computer algorithms so that the analysis can be performed on more than just a few variables (Vermunt & Magidson, 2002). LCA has several advantages over traditional multivariate techniques. It does not rely on traditional assumptions such as linearity, normality, or homogeneity of variance – which are often violated in practice – and is designed to classify cases into latent classes based on categorical survey items (Vermunt

& Magidson, 2002). In LCA, cases are classified into clusters based on maximum likelihood estimates and the Bayesian methodology. It begins with an equal prior probability (for the basic LCA model) for each case being classified into each latent class. Then, with the observed responses to the manifest variables, posterior probabilities are estimated and are used to indicate the probability of an observed case being classified into a given class (Hagenaars & McCutcheon, 2002). It also provides parameter estimates and goodness-of-fit indices that measure model fit.

Using SAS PROC LCA (Version 1.2.7; 2011), models with increasing numbers of latent classes (from two to eight classes) were tested. Different models were then compared using goodness-of-fit indices including the likelihood ratio G^2, Akaike Information Criterion, Bayesian Information Criterion, entropy R^2, and the number of parameters estimated. Balancing between model fit, parsimony, and interpretability, the most plausible solution was selected. In the analysis for research question one, the association between this student profile variable and senior students' civic engagement scores was further explored after adjusting for students' civic engagement scores in their freshman year, as well as other student characteristics.

3.2.3 Institution-level Variables

Variables at the institution-level came from two sources. HERI collects information such as institution type (i.e., four-year university, four-year college, and two-year college), institution control (i.e., public, private), and selectivity (i.e., median SAT total scores (sum of Verbal and Math scores) or the equivalent ACT composite scores).

Table 3.6 Descriptions of Institution-level Variables

Variables	Descriptions
Institution Control	Dichotomous variable: 1=Private, 0=Public
Institution Type	Dichotomous variable: 1=Four-year universities, 0= Four-year colleges
Institutional Selectivity	Continuous variable: Median of SAT Math and Verbal scores or ACT composite scores that were put on the SAT score scale
Institutional Stratification*	Four-category variable: Public universities and colleges, Private/nonsectarian 4yr colleges, Catholic 4yr colleges, and Other religious 4yr colleges
Office of Education Region*	Six-category variable: New England, Mid-Atlantic, Southeast, Great Lakes, West, and Far West
Basic Carnegie Classification*	Three-category variable: Baccalaureate colleges, Master's colleges and universities, and Research universities
Undergraduate Instructional Program*	Five-category variable: Arts & Sciences focus, Arts & Sciences focus plus professions, Balanced Arts & Sciences, Professions plus Arts & Sciences, and Others
Enrollment Profile*	Four-category variable: Exclusively undergraduate, Very high undergraduate, High undergraduate, and Others
Size and Setting*	Four- category variable: Small four-year, Medium four-year, Large four-year, and Others
Degree of Urbanization (Urban-centric locale)*	Four-category variable: City (large, midsize, and small), Suburb (large, midsize, and small), Town (fringe, distant, and remote), and Rural (fringe, distant, and remote)

* To protect institutions' confidentiality and due to the availability of the institutions in the datasets, these IPEDS/CIRP variables are recoded in the study. The new categories are different from those original IPEDS/CIRP categories.

IPEDS provides more detailed information such as the Carnegie classifications and the degree of urbanization (urban-centric locale) of the institution. It is worth noting that, to protect institutions' confidentiality, some IPEDS or CIRP variables were recoded with combined categories when the number of institutions was small in the original IPEDS/CIRP category. For example, the basic Carnegie classification variable from IPEDS originally has 35 categories. But institutions from many categories were not available in the CIRP sample; for some categories, the classifications are too detailed and include too few of the institutions. After recoding, the basic Carnegie classification variable in this study has three categories – Baccalaureate colleges, Master's colleges and universities, and Research universities. A description of these institution-level variables is presented in Table 3.6.

Since the measure of school value-added is more valuable in comparing peer institutions rather than ranking schools overall, these school-level variables (such as the institution's Carnegie classification and selectivity) were not directly used in the value-added models. Rather, they were used to categorize institutions into more homogeneous groups to make valuable comparisons. In other words, value-added analyses first were carried out across all available institutions to obtain the value-added estimates for each institution. Then, these estimates were restandardized within each peer comparison group of institutions.

3.2.4 Other Variables

Since the students participating in the CIRP surveys were not a true random sample, but a voluntary, self-selected sample, post-stratification weights were created in this study to reduce biased estimates due to nonrepresentativeness in the voluntary survey sample. Post-stratification is a common practice in survey analysis. It aims at matching the distribution of respondents in the sample grouped by strata with that of the population (Little, 1993). In this study, two strata – gender and race/ethnicity – are chosen to calculate the post-stratification weights within each institution. The sample information in each post-stratum (e.g., percentage of white male students who responded to CIRP in an institution) comes from the CIRP surveys and the corresponding population information comes from the IPEDS data. However, nonresponse bias cannot be completely removed and the sample estimates will be biased if post-stratification is not properly used (Zhang, 1999). In this study, the differences in the percentages of demographic subgroups between the institutional population and the final sample were tested. Decisions of using post-stratification were based on whether significant differences exist between sample and population properties.

Some other variables were used in the data preparation process, but were not included in the value-added models. These variables are generally related to students' undergraduate status and college transfer information. They are:

- In what year did you graduate from high school?
- Please indicate your enrollment status (full-time/part-time).

- Please indicate the highest degree you will have earned as of [the end of this school year].
- What year did you first enter your first college?
- What year did you first enter this college?
- Have you transferred from another college?
- Did you enter college more than four years ago?

3.3 Analytical Strategy

3.3.1 Preliminary Descriptive Analyses

The preliminary descriptive analyses section includes results at both the student level and the institution level. First, this section reports results from the creation of the student-level outcome variable, the predictor variable, and other variables (e.g., the post-stratification weight). Specifically, for the three-stage study on the civic engagement scale, the reported results include those from each analysis stage and the distribution of civic engagement in different CIRP survey administrations. Moreover, the distribution of the student profile variable, the construction of the post-stratification variable, and cross-tabulations between these constructed variables and students' civic engagement level are summarized.

Next, the preliminary descriptive analyses include (1) a summary of the institutional characteristics such as institution type, control, selectivity, and Carnegie classifications; (2) a comparison of the institutional mean civic engagement scores among freshman and senior students and the changes in civic engagement scores in different

institutional subgroups; and (3) the grouping of institutions into peer groups based on the institutions available and their characteristics.

3.3.2 Research Question One

The first research question asks: To what extent are undergraduate students' characteristics associated with a measure of civic engagement in the senior year after adjusting for their level of civic engagement as freshmen? This research question aims at identifying student covariates that are associated with civic engagement and can be used in later models to estimate institutional value-added. Because of the multilevel structure of the data with students nested within institutions, two-level hierarchical linear models were used to account for the dependency among students from the same institution (Raudenbush & Bryk, 2002). The models were carried out in stages to predict students' civic engagement in CIRP senior surveys using the longitudinal dataset. The statistical models generally can be expressed as

Level 1 (Student level):

$$Civic_{ij,S} = \beta_{0j} + \beta_{1j}(Civic_{ij,F}) + \beta_{2j} X_{2ij} + \ldots + \beta_{pj} X_{pij} + r_{ij}$$

Level 2 (Institution level):

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

$$\ldots$$

$$\beta_{pj} = \gamma_{p0} + u_{pj} \qquad (3.2)$$

where *i* denotes students within schools, and *j* denotes institutions;

$Civic_{ij,S}$ is the measure of civic engagement for student *i* in institution *j* in their senior year;

$Civic_{ij,F}$ is the measure of civic engagement for student *i* in institution *j* in their freshman year;

$X_{2ij} \ldots X_{pij}$ are *p*-1 student-level covariates for student *i* in institution *j*;

β_{0j} is the grand mean of the civic measure for institution *j*, adjusted for the covariates $Civic_{ij,F}$, X_2, …, X_p;

β_{1j}, …, β_{pj} are the regression coefficients for institution *j*, associated with the covariates $Civic_{ij,F}$, X_2, …, X_p;

r_{ij} is the random error (or residual) in the level 1 equation, where $r_{ij} \sim N(0, \sigma^2)$ and σ^2 is the variance of the student-level residuals;

γ_{00} is the intercept for the level 2 equation which is the grand mean of the civic measure across all institutions;

γ_{10}, …, γ_{p0} are constants representing the common values of the *p* regression coefficients across all institutions; and

u_{0j}, …, u_{pj} are random effects in the level 2 equations, where $u_{pj} \sim N(0, \tau_p)$ and τ_p is the variance of the institution-level residuals.

With the adoption of multiple imputation technique, five datasets were generated where each has one set of imputed values on missing responses. HLM 6.08 software was

used to analyze the multiply imputed datasets and obtain the final results synthesized from separate estimates on each available dataset.

The first research question focuses on the association between student-level variables and students' civic engagement scores in the senior year after adjusting for civic engagement scores in the freshman year, so institutional-level variables were not included. During the model building procedure, each model was carried out in stages. In the first stage, an unconditional model was built. An unconditional model includes no predictor at either level 1 or level 2. The model allows the partition of the total variability in civic engagement into within- and between-institution variance components, providing baseline information on variance decomposition.

In the second stage, adjusted civic engagement scores of the same students from their freshman year were added to the unconditional model. Since measurement error in an independent variable in a regression model can result in attenuation bias in the estimation of the regression slope (Fuller, 1987; Shadish, Cook, & Campbell, 2002), civic engagement scores in the freshman year were entered into the model after adjusting for measurement error.[14] This measurement error adjustment method has been successfully used in the literature, such as in O'Dwyer, Carey, and Kleiman (2007). It is hypothesized that the same student's civic scores between freshman year and senior year are highly correlated, therefore, this model should explain a large percent of the available variance in the civic outcome.

[14] The formula for this adjustment is $X_{adj} = \overline{X} + r(X - \overline{X})$, where *r* is the reliability of freshmen's civic engagement score, or Cronbach's alpha.

In the third stage, other student-level covariates (e.g., gender, race, religion, college GPA, SAT/ACT, SES, educational aspiration, student profile) were added by group to explore the association between civic engagement and students' characteristics. Specifically, four groups of student-level variables were entered into the model following the sequence of student demographics and family background, high school educational outcomes, college educational outcomes and experiences, and student profile. Because the goal is to identify student-level variables that are associated with senior students' civic engagement, a less stringent significance level was used at the beginning to select a larger group of covariates with potential associations. For each group of covariates, those statistically significant at the .05 level were retained. After examining four groups of variables separately, all retained variables were added to the next model. Here, a more stringent criterion, .01 significance level, was used to select variables in the final model at stage three. At the end of this stage, the relationships between all the retained level-1 predictors and the civic outcome were examined across institutions to determine whether the slope associated with each covariate should be fixed or allowed to vary. If there was no significant variation in the level-1 slopes across institutions, the level-1 slopes were fixed. Otherwise, the level-1 slopes were allowed to vary. However, because the number of units at level 2 is much smaller than that at level 1, the slope estimates can be far less reliable. As suggested from past experiences (Raudenbush & Bryk, 2002), a slope would

be fixed when the reliability[15] of a random level-1 coefficient drops below 0.05 (see Raudenbush & Bryk, 2002, p.125).

In the final stage, interactions between level-1 covariates were tested to explore potential interaction effects. Similarly, level-1 slopes associated with significant interaction terms were tested to determine whether the slopes should be allowed to vary or not. Therefore, the final model was different from equation 3.2 above in terms of certain fixed level-1 slopes.

Different centering decisions (i.e., group-mean centering, grand-mean centering, and raw metric without centering) result in fundamental differences in the interpretation of estimated intercept and slope parameters in multilevel models. Based on recommendations from empirical analyses, separate centering decisions were made for each research question in this study. These centering decisions are introduced in section 3.3.5, after representing the basic analytical strategies for all research questions.

3.3.3 Research Question Two

The second research question asks: To what extent are institutional characteristics (such as school type, selectivity, location, and type of location) associated with a measure of senior students' civic engagement after adjusting for their level of civic engagement as freshmen, as well as their characteristics? To answer this question, new HLMs were built from the final model in research question one (i.e., the model controlling for civic

---

[15] The reliability of level-1 slopes are estimated as: reliability($\hat{\beta}_p$) = $\frac{1}{J}\sum_{j=1}^{J}\tau_{pp}/(\tau_{pp}+v_{ppj})$ for each p = 1, …, P, where τ_{pp} is the parameter variance and v_{ppj} is the error variance (Raudenbush & Bryk, 2002).

engagement at the freshman year and other student-level covariates). These statistical models take the general form of:

Level 1 (Student level):

$$Civic_{ij,S} = \beta_{0j} + \beta_{1j}(Civic_{ij,F}) + \beta_{2j} X_{2ij} + \ldots + \beta_{pj} X_{pij} + r_{ij}$$

Level 2 (Institution level):

$$\beta_{0j} = \gamma_{00} + \gamma_{01} W_{1j} + \ldots + \gamma_{0q} W_{qj} + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} W_{1j} + \ldots + \gamma_{1q} W_{qj} + u_{1j}$$

$$\ldots$$

$$\beta_{pj} = \gamma_{p0} + \gamma_{p1} W_{1j} + \ldots + \gamma_{pq} W_{qj} + u_{pj} \qquad (3.3)$$

where $W_{1j}, \ldots, W_{qj}$ are q institution-level covariates for institution j; and $\gamma_{01}, \ldots, \gamma_{pq}$ are the regression coefficients associated with the institution level covariates $W_{1j}, \ldots, W_{qj}$.

Comparing equations 3.2 and 3.3, the only difference is that institutional characteristics were added at level-2. Institution level covariates $W_{1j}, \ldots, W_{qj}$ were added to explore the associations between the level-1 intercept (i.e., adjusted institutional mean), level-1 slopes, and institutional characteristics. The model building process involves the following new stages: Institutional covariates first were added one by one at level-2 to predict the intercept of level-1. When the covariate was not significant, it was removed from the model. In the following stage, when a level-1 slope was allowed to vary across institutions, institution-level covariates were added to explain the variance in the level-1 slopes. Finally, cross-level interactions (e.g., the interaction term of race/ethnicity and

institutional type) were examined on whether a significant amount of variance can be explained by the interaction term and whether the level-1 slope can be reliably estimated in the model. It is worth noting that the HLM model in the above equation represents a full model which takes into account all possible covariates and interactions (i.e., an intercepts- and slopes-as-outcomes model). The final model took a much simpler form than what is presented here.

3.3.4 Research Question Three

The third research question focuses on value-added methodologies. It asks about the similarities and differences among the results of different models in estimating institutional effectiveness. Institutional value-added scores were first calculated using three different estimation methods. Next, based on institutional characteristics (such as the Carnegie classifications, institutional type, selectivity, and type of location), schools with similar background were selected to form homogeneous peer groups. Value-added estimates then were calculated and the relative positions of institutions within the same peer group were compared across different methods.

As described in the previous chapter (see section 2.3.6), Method 1 is based on two OLS regression models and the value-added score is the difference in differences of the observed and predicted civic measures. Method 1 takes the statistical form of:

$$\overline{Civic}_{j,F} = \alpha_F + \beta_F(\overline{X}_{j,F}) + u_{j,F} \qquad (j = 1,\ldots, n) \qquad (3.4)$$

$$\overline{Civic}_{j,S} = \alpha_S + \beta_S(\overline{X}_{j,S}) + u_{j,S} \qquad (j = 1,\ldots, n) \qquad (3.5)$$

where *F* denotes freshman students, *S* denotes senior students, and *j* denotes institution, which is the unit of analysis;

$\overline{Civic}_{j,F}$ is the mean civic outcome among freshman students in institution *j*;

$\overline{Civic}_{j,S}$ is the mean civic outcome among senior students in institution *j*;

$\overline{X}_{j,F}$ is the mean of a student-level covariate among freshman students in institution *j*;

$\overline{X}_{j,S}$ is the mean of a student-level covariate among senior students in institution *j*;

α_F is the intercept of the freshman regression model;

α_S is the intercept of the senior regression model;

β_F is the slope of the freshman regression model associated with the student-level covariate;

β_S is the slope of the senior regression model associated with the student-level covariate;

$u_{j,F}$ is the residual between the observed and expected civic mean scores from the freshman regression model for institution *j*; and

$u_{j,S}$ is the residual between the observed and expected civic mean scores from the senior regression model for institution *j*.

The institutional value-added score is $u_{j,S} - u_{j,F}$, the difference score between the senior and freshman residuals of that institution. The association between noncognitive variables and students' academic achievement such as SAT has been found to be weak (see Knox et al., 1993; Robbins et al., 2006), so other student-level covariates may emerge from the analysis of research question one. Based on the findings from research

question one, three sets of student covariates were selected to include in the above equations.

Because some VAMs in the K-12 setting require multiple years of data to obtain a precise estimate on school effectiveness (e.g., the SAS EVAAS model), the above approach was replicated two times with data from different years. Institutional value-added scores in 2002 were estimated using both freshman and senior surveys conducted in 2002 (i.e., TFS2002 and CSS2002); institutional value-added scores in 2006 were estimated using both freshman and senior surveys conducted in 2006 (i.e., TFS2006 and CSS2006).

Note that the departure from OLS model's assumptions such as linearity and homoscedasticity of the error variances in any of the freshman and senior regression models may affect the accuracy of the institutional value-added estimates. Before carrying out each regression model, the linearity, homoscadasticity, and normality assumptions were tested. Similarly, homogeneity of level-1 variances and model fit were also examined before running the HLMs in Method 2 and Method 3 below.

Method 2 adopted an HLM-based approach and adjusted for a proxy of students' prior civic measure. Similar to equation 2.3, this two-level HLM can be expressed as:

Level 1 (Student Level):

$$Civic_{ij,S} = \beta_{0j} + \beta_{1j}(X_{ij,S} - \overline{X}_{j,S}) + r_{ij}$$

Level 2 (School Level):

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\overline{X}_{j,S}) + \gamma_{02}(\overline{Civic}_{j,F}) + u_{0j}$$

$$\beta_{1j} = \gamma_{10} \qquad (3.6)$$

Here, $Civic_{ij,S}$ is the civic score of senior student *i* in institution *j* and $\overline{Civic}_{j,F}$ is the mean civic score among freshman students in institution *j*. These seniors and freshmen are two student cohorts who studied in the same institution. X is the same student-level covariate used in Method 1 with the use of group-mean centering. Similar to the HLM model adopted by CLA (equation 2.3), γ_{10} is a constant representing the regression coefficient associated with the student-level covariate. In other words, the level-1 regression slope is assumed to be the same for all institutions.[16] Institutional value-added score is the level-2 residual u_{0j}. In other words, the value-added estimate is the difference between observed and expected civic engagement measures among senior students after adjusting for a proxy of freshmen's civic measure in that institution and other student characteristics. Just as with Method 1, Method 2 was also replicated twice with CIRP survey data from 2002 and 2006.

[16] In later analyses, $\beta_{1j} = \gamma_{10} + u_{1j}$ was tested. But no significant random effect was found for the level-1 slope.

Next, Method 3 employed longitudinal data that contained information on the same student at two time points – once in freshman year and once in senior year. The statistical equations are:

Level 1 (Student Level):

$$Civic_{ij,S} = \beta_{0j} + \beta_{1j}(X_{ij} - \overline{X}_j) + \beta_{2j}(Civic_{ij,F}) + r_{ij}$$

Level 2 (School Level):

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\overline{X}_j) + u_{0j}$$

$$\beta_{1j} = \gamma_{10}$$

$$\beta_{2j} = \gamma_{20} \qquad (3.7)$$

Here, $Civic_{ij,S}$ is the civic score of student *i* in institution *j* measured in his/her senior year. $Civic_{ij,F}$ is the civic score of the same student *i* in institution *j* measured in his/her freshman year. X is the same student-level covariate used in Methods 1 and 2. γ_{10} is a constant representing the regression coefficient associated with the student-level covariate *X*. γ_{20} is a constant representing the regression coefficient associated with the civic engagement score at the freshman year.[17] Institutional value-added score is the level-2 residual u_{0j}, the difference between observed and expected civic engagement measures among senior students after adjusting for those students' civic engagement

[17] In later analyses, $\beta_{2j} = \gamma_{20} + u_{2j}$ was tested. Although u_{2j} was marginally significant (p = .03), its variance component was only 0.003. Moreover, a likelihood-ratio test was conducted to compare the deviance statistics between equation 3.7 and an alternative model with the addition of a random slope for $Civic_{ij,F}$. The addition of the random slope was negligible ($\chi^2_{(2)} = 3.75$, p = .15). Therefore, the level-1 slope β_{2j} was fixed.

scores in their freshman year and other student characteristics in each institution. Method 3 used the longitudinal data representing matched student responses from TFS2002 and CSS2006.

Finally, institutional value-added scores estimated from the three methods were compared. The purpose of this study is not to determine which method is superior but to provide empirical evidence of the similarities and differences in estimating institutional effectiveness with regard to civic engagement. Within-method stability and between-method agreement were the focus of the comparison. Since different models based on different years of data were carried out using Methods 1 and 2, within-method stability was first examined for these two methods. Specifically, within-method stability was indicated by the Pearson correlation coefficient between the value-added estimates based on different years of data. These models were further examined among institutions with a large number of respondents to explore the role of sample size in estimating institutional value-added scores. Next, results based on the longitudinal method were reported. All three methods were then investigated for between-method agreement (and differences). Correlations and scatterplots were used to compare different pairs of models. Institutional rankings were also categorized into quartile groups and the changes in the quartile groups in different methods are discussed. The three models estimated institutional value-added in the same year, 2006, were selected and compared. This comparison revealed some differences between the OLS-based and HLM-based cross-sectional approaches, which are currently used in VSA, and an HLM-based longitudinal approach. In the end, institutions' relative positions in each peer group were visually

presented in graphs. This also revealed some differences between the longitudinal and cross-sectional value-added methods.

3.3.5 Centering Decisions

As mentioned earlier, estimated intercept and slope parameters in multilevel models are interpreted differently when using different centering decisions (i.e., group-mean centering, grand-mean centering, and raw metric without centering). However, centering decisions are often overlooked or misunderstood in practice (Enders & Tofighi, 2007; Hofmann & Gavin, 1998; Kreft, de Leeuw, J., & Aiken, 1995; Raudenbush, 1989; Raudenbush & Bryk, 2002). Thus, this section focuses on centering decisions made to answer the research questions in the study where two-level HLMs are used.

Researchers have provided practical recommendations on the choice of centering through empirical analyses. In short, Raudenbush and Bryk (2002) recommended using group-mean centering when estimating person-level effect because it provides an unbiased estimate of the within group regression coefficient (β_w). However, in organizational research applications with the purpose of estimating level-2 effects while adjusting for level-1 covariates, grand-mean centering is more appropriate. The reason is that, after grand-mean centering level-1 covariates, the level-1 intercept (β_{0j}) is the adjusted mean across level-2 units; the estimate of level-2 effect "will be adjusted for differences between organizations in … the level-1 explanatory variable" (p.142).

Enders and Tofighi (2007) also provided advice in making centering decisions on interaction terms. Group-mean centering is recommended when investigating cross-level

interactions (i.e., interaction between level-1 and level-2 variables) and interactions between level-1 variables. Grand-mean centering is recommended for investigating interactions between level-2 variables. In addition, the same concept can be applied to centering binary (i.e., dummy-coded) variables. Although it may seem unnecessary, a consistent centering method across continuous and binary variables produces a straightforward interpretation on estimated parameters.

Applying the above suggestions in the three research questions in this study, group-mean centering was used in the first research question of this study when examining the association between senior students' civic engagement scores and student-level variables and exploring the interactions between student-level variables. Group-mean centering was used on all covariates at the student-level, including continuous variables such as SAT/ACT scores and binary variables such as gender. Group-mean centering student-level covariates provide more accurate estimates on student-level slopes and a clear decomposition of the within- and between-school effects. However, to account for the institutional mean differences on those student covariates, those institutional means were included at level-2. Centering decisions were less complex at level-2 because only two options are available – raw metric and grand-mean centering. Since all the institutional mean variables are continuous, such as institutional mean of freshmen's civic engagement scores, grand-mean centering was used so that the value of zero is meaningful in interpretations.

As for research question two, centering decisions were also made separately for variables at different levels. At the student-level, grand-mean centering was used when

the purposes were to estimate institution-level effects while adjusting for student-level covariates and exploring interactions between institution-level variables. Yet, group-mean centering was used in examining cross-level interactions. At the institutional-level, grand-mean centering was used for continuous variables, such as institutional selectivity. For the ease of interpretation, dichotomous variables at level-2 retained their raw metrics.

Centering decisions on the three methods in research question three were also considered separately. Since Method 1 adopted OLS regression and had institution as the unit of analysis, all variables involved are institutional means. Method 2 followed the same centering decision as VSA. As shown in equation 3.6, group-mean centering was used at the student-level and grand mean-centering was used at the institution-level. For Method 3, as the goal is to estimate institutional effect (or organizational research in Raudenbush and Bryk's terms), the same centering decisions were made as those in research question two. Based on these centering decisions, appropriate interpretations are made in the results in chapter 4.

CHAPTER 4. RESULTS

This chapter presents results from the analyses outlined in chapter three. It is organized into four sections. The first section reports results from descriptive analyses, which includes a three-stage study on civic engagement and the creation of some student level variables such as student profile and post-stratification weights. After variable construction, the first section further reports the distribution of the civic engagement measure in four different CIRP survey administrations, students' changes in civic engagement from freshman to senior year, a summary of the institutional characteristics, the grouping of institutions into peer groups, and cross-tabulations between the constructed variables and other student or institution-level background variables.

Each of the next three sections answers one of the research questions of this study. In sections two and three, hierarchical linear models were carried out in stages and the results are reported at the end of each stage. Because some institutions had a relatively small number of respondents, to obtain more reliable estimates, the models were first built among the 90 institutions that had at least 20 freshmen and 20 seniors responding to the surveys. In section two[18], the models were also replicated among institutions that had at least 100 freshmen and 100 seniors responding to the surveys. Finally, section four reports the results from three value-added methods – OLS-based cross-sectional method, HLM-based cross-sectional method, and HLM-based longitudinal method. Institutional value-added scores are estimated and results are reported for each method. Since

[18] Due to the small number of institutions available (n=26), the models were not replicated in section three when several institutional-level covariates were added to the model.

different models based on different years of data were built using Methods 1 and 2, within-method stability is reported for these cross-sectional methods. After presenting results from Method 3, between-method agreement across all three methods and an indicator of precision provided by hierarchical linear models are further examined. In the end, institutional value-added estimates are compared within nine peer groups across the three methods used.

4.1 Descriptive Analyses

4.1.1 Civic Engagement

Intending to minimize measurement error, a three-stage study was carried out to derive the indicator for students' civic engagement. Results from each of these three stages are reported below.

4.1.1.1 Stage One: Exploratory Factor Analysis and Unidimensionality

Previous studies have found that civic engagement is a multidimensional construct (Keeter, Zukin, Andolina, & Jenkins, 2002; Schulz et al., 2010). In this study, 17 items from the CIRP surveys were first selected because they satisfy two criteria: (1) They are common items measured in all four survey administrations (i.e., TFS2002, TFS2006, CSS2002, and CSS2006); and (2) The items ask about students' attitudes, beliefs, and past activities related to civic engagement. Next, these items were analyzed in a series of principal component analyses in each of the four survey administrations. In the end, a consistent two-factor solution across all surveys administrations emerged.

Table 4.1 shows the results of the factor analyses, the reliability analyses on each derived factor, and total sample sizes used in each analysis.

Table 4.1 Results of Exploratory Factor Analyses and Reliability Analysis on Civic Engagement Related Items

Item Content[a]	Factor Loadings[b]							
	TFS2002		TFS2006		CSS2002		CSS2006	
	1	2	1	2	1	2	1	2
The importance of								
1. Participating in a community action program	.773		.782		.804		.807	
2. Promoting racial understanding	.717		.733		.744		.746	
3. Becoming a community leader	.719		.710		.732		.729	
4. Influencing social values	.642		.651		.687		.688	
5. Keeping up to date with political affairs	.615		.629		.631		.635	
6. Developing a meaningful philosophy of life	.622		.620		.608		.617	
7. Becoming involved in programs to clean up the environment	.610		.627		.589		.604	
8. Helping others who are in difficulty	.582		.591		.607		.610	
Time spent on								
9. Student clubs/groups		.643		.621		.704		.723
How often did you:								
10. Perform volunteer work		.615		.614		.532		.492
11. Vote in a student election		.567		.591		.690		.690
Total variance explained (%)	34.64	11.86	35.03	11.73	36.24	12.98	36.73	12.84
Scale reliability	0.83	0.46	0.84	0.43	0.84	0.54	0.84	0.52
Total sample size	339,751		312,798		32,873		27,043	

Extraction Method: Principal Component Analysis.

[a] Items 1-8 are on a 4-point scale: 1=Not important; 2=Somewhat Important; 3=Very Important; 4=Essential. Item 9 is on an 3-point scale: 1=None; 2=Less than 10 hrs; 3=More than 10 hrs. Items 10-11 are on a 3-point scale: 1=Not at all; 2=Occasionally; 3=Frequently.

[b] Factor loadings that are less than .4 are not shown.

All items in Table 4.1 have relatively high loadings (ranging from .492 to .807) on the construct they measure. The first factor is related to students' attitudes toward civic engagement and explains 35-37% of the total variance, while the second factor reflects students' participation in civic related activities and explains 12-13% of the total variance.

Further factor analyses on the items within each factor also confirmed that all the extracted factors are unidimensional. The first derived factor consists of eight items and has a high reliability that is above 0.8. The second derived factor consists of only three items and has a relatively low reliability that ranges from 0.43 to 0.54. The first derived factor also conforms to findings from previous studies of civic engagement using the CIRP data (e.g., Astin & Sax, 1998; Pascarella et al., 1988). Yet, considering the low reliability of the second factor, the three-item scale was not included as part of the civic engagement dependent measure in this study. The sum of these three items was used only to reflect students' activities involving volunteer work, voting in student elections, and involvement in student clubs/groups during college in later regression analysis. The bottom row of Table 4.1 listed the sample sizes used in these exploratory factor analyses, which were relatively large.

4.1.1.2 Stage Two: Confirmatory Factor Analysis and Multisample Confirmatory Factor Analysis

As outlined in chapter three, CFA was conducted first within each of the four CIRP administrations to test the factor structure of civic engagement. To test the

structure consistency across administrations, multisample CFA was carried out next to examine whether the eight items in the CIRP survey measure the same construct in different survey administrations. In other words, multisample CFA was used to investigate whether the factor structure of the civic engagement scale derived from the CIRP surveys was invariant across four student samples with different ages and different grade levels. Since the widely used fit indices such as χ^2 can be affected by sample size with the tendency of rejecting good models with large sample sizes, 500 cases were randomly selected from each CIRP administration. The individual sample CFA models were estimated from the randomly selected cases, but the polychoric correlation and asymptotic covariance matrices were calculated from the entire sample to obtain robust correlation and covariance estimations. Accordingly, the multisample CFA analysis included 2,000 cases in total.

The goodness-of-fit statistics of the individual sample CFA are summarized in Table 4.2 (see Figure 4.1 for a conceptual path diagram of the single sample CFA). Although the minimum fit function chi-square tests were rejected (here, the *p*-values were expected to be greater than 0.05), standardized root mean square residuals (SRMRs) were below 0.08 in all four CIRP administrations; the RMSEA values were below 0.06 in three CIRP datasets, with CSS02 being the only exception; and the TLI values reached 0.95 in the two freshman surveys. These fit statistics indicate a good fit for the 8-item measurement model to the CIRP freshman data and a moderate fit to the senior survey data.

Table 4.2 Goodness-of-Fit Statistics of the Individual Sample CFA for Eight CIRP Items

	TFS02	TFS06	CSS02	CSS06
Minimum fit function chi-square (degrees of freedom)	51.61 (20)	49.90 (20)	59.94 (20)	49.19 (20)
p-value	0.00013	0.00023	0.00001	0.00029
RMSEA	0.056	0.055	0.063	0.054
TLI	0.95	0.95	0.92	0.94
SRMR	0.06	0.06	0.07	0.06

Note. RMSEA = root mean square error of approximation; TLI = Tucker-Lewis index; SRMR = standardized root mean square residual.

Next, multisample CFA was used to examine measurement invariance across different CIRP survey administrations. Jöreskog and Sörbom (2001) have suggested multiple psychometric elements to test in multisample CFA. Two of the elements, factor pattern invariance and factor loading invariance, are the most fundamental for establishing conceptual and metric equivalence and were tested in this study. The assessment of invariance involved testing the difference between the baseline model that was factor pattern invariant and a second model that was factor loading invariant. A chi-square difference test was used to examine whether the factor loadings of the measurement model were invariant across the four student samples. Figure 4.1 is the path diagram for the baseline model. The RMSEA index is below 0.06, which indicates a good fit by the 8-item measurement model to the data.

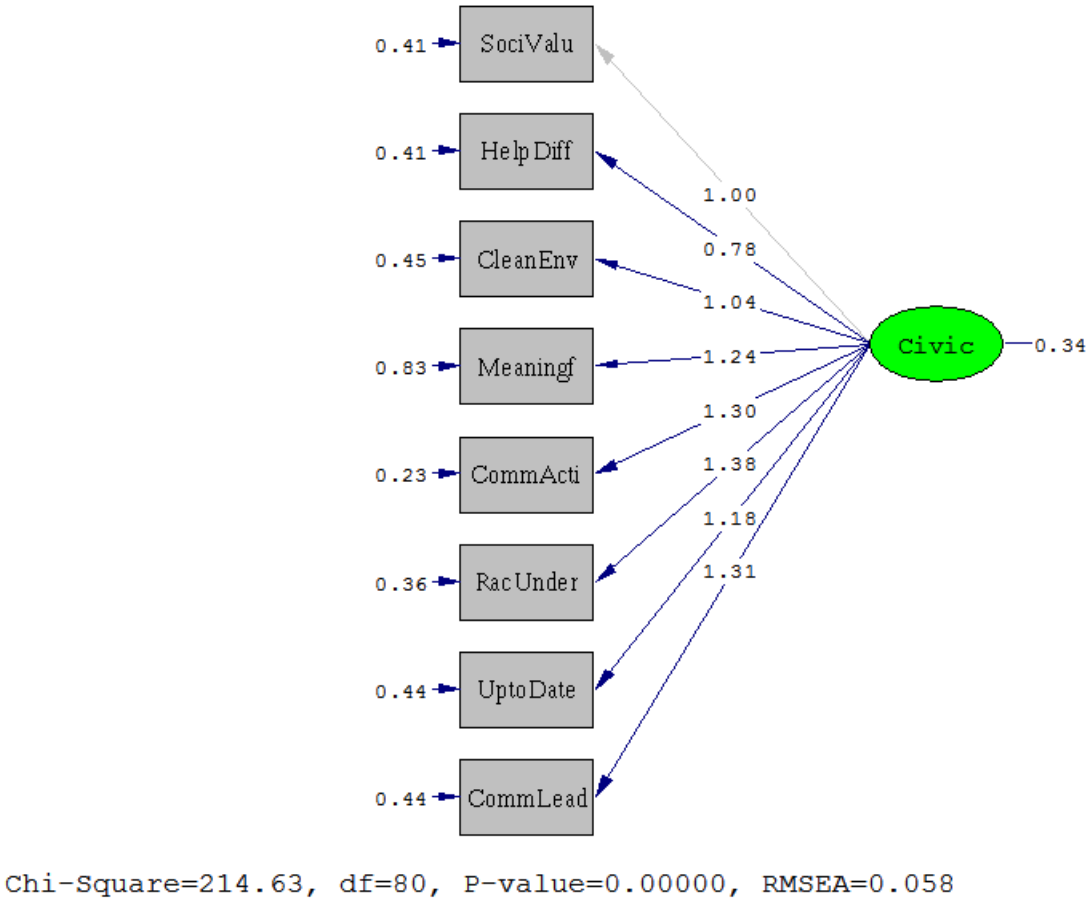


Figure 4.1 Path Diagram for the Multisample CFA Baseline Model

Table 4.3 shows the results of the factor pattern invariance model, the factor loading invariance model, and the measurement invariance test across four survey administrations. First, the fit indices of the models indicate that both measurement models fit the data well; while the factor loading invariance model fits the data slightly better, as seen from the smaller RMSEA and the larger TLI and SRMR indices. The chi-square difference test between the two models has a *p*-value of 0.87. The large *p*-value suggests that there is sufficient evidence that the null hypothesis should be retained. In

other words, factor loadings of the civic engagement measurement model were invariant across the four student samples.

This chi-square test result also conformed with findings in the comparative fit index (CFI) reported in Table 4.3. In a single sample CFA, a CFI value of 0.90 and above indicates a good fitting model. The changes in CFI is recommended for use as a supplement to other fit indices because it is independent of sample size and model complexity (Cheung & Rensvold, 2002; Vandenberg & Lance, 2000). The suggested criterion is that changes in CFI of 0.01 or less indicate retaining the null hypothesis of measurement invariance. The change in CFI is 0.002 in Table 4.3. It confirms the previous conclusion that factor loadings were not statistically different across the four CIRP survey administrations. This 8-item civic engagement model then moved to the next stage.

Table 4.3 Results of Measurement Invariance in Multisample CFA

Measurement Model	Chi-Square (*df*)	RMSEA	TLI	SRMR	CFI
Baseline Model & Factor Pattern Invariance	214.63 (80)	0.058	0.94	0.06	0.957
Factor Loading Invariance	228.62 (101)	0.050	0.96	0.08	0.959
Difference	13.99 (21) [ns]				0.002

Note. RMSEA = root mean square error of approximation; TLI = Tucker-Lewis index; SRMR = standardized root mean square residual; CFI = Comparative Fit Index.
[ns] $p = .87$

4.1.1.3 Stage Three: Item Response Theory (IRT) Scaling Analysis

The 1-parameter IRT (Rasch) scaling analysis was conducted using Winsteps software (Lincare, 2006). The results of the Rasch item calibration analysis show an item reliability of 1.00 and a person reliability of 0.79, indicating good reproducibility properties for the 8-item civic engagement scale. Table 4.4 lists the item difficulty parameter (or the location of the item on the latent trait continuum), the step parameter, and item fit statistics for the civic engagement scale. For ease of reading, the items are abbreviated and maintain the same sequence as Table 4.1. The column labeled "Location" shows item difficulty parameters, which were rescaled with a mean of 50 and standard deviation of 10. The smaller this parameter is, the easier this item is for students to endorse. The table shows that the easiest item to endorse is the importance of helping others who are in difficulty. The hardest item to endorse is the importance of becoming involved in programs to clean up the environment. Since all eight items have a 4-category response option ranging from "Not Important" to "Essential," three step difficulty parameters were estimated, which shows the difference in endorsing adjacent categories. Table 4.4 shows that the step parameters for all eight items range are sufficiently large and in sequential order. The values of the step parameters were different across different items which means that students adopt different criteria for the four response categories for different items.

Moreover, the INFIT and OUTFIT statistics are the two most commonly used goodness of fit indices for the Rasch model. INFIT and OUTFIT values near 1 are desirable. Specifically, those ranges from 0.7 to 1.3 show a good fit between the item

and the partial credit model, which also indicate adequate construct homogeneity of the scale or scale unidimensionality (Wright & Lincare, 1994). Table 4.4 shows that the eight items' INFIT and OUTFIT values range from 0.75 to 1.14, which indicate a good model-data fit.

For a more detailed examination of item difficulty and student ability, Appendix B reports the Rasch item maps (both with and without expected score zones) and category probability curves for each item. They show that the items and the students are approximately normally distributed; the item difficulties are appropriately spread along the latent continuum; and item thresholds are in sequential order with suitable distances between adjacent item categories.

Table 4.4 Item Parameters and Fit Statistics for the 8 Item Civic Engagement Scale

Item	Location	Step 1	Step 2	Step 3	INFIT Mean Square	OUTFIT Mean Square
1. CommActi	55.63	-19.93	3.32	16.61	0.77	0.75
2. RacUnder	52.02	-18.99	2.93	16.06	0.85	0.85
3. CommLead	52.49	-17.64	2.40	15.24	0.95	0.95
4. SociValu	48.39	-20.62	1.19	19.43	1.02	1.03
5. UptoDate	51.38	-18.08	1.81	16.27	1.07	1.10
6. Meaningf	46.60	-14.32	1.10	13.22	1.12	1.13
7. CleanEnv	59.05	-19.83	4.74	15.10	1.10	1.07
8. HelpDiff	34.44	-26.16	3.12	23.04	1.13	1.14

After item calibration, the second step of the Rasch analysis was to generate student scores on the civic engagement scale for all cross-sectional and longitudinal

datasets used in this study. In these Rasch analysis, item parameters were anchored at the estimated values from step one. Table 4.5 summarizes some descriptive statistics of civic engagement scores in each of the four CIRP survey administrations. All first-time full-time undergraduate students who responded to at least six of the eight items on the civic engagement scale are included in the table. Table 4.5 shows that the civic engagement scores range between -8 and 101 among all respondents. The mean civic engagement scores are higher among senior students than those of freshman students. Moreover, the variances of civic engagement scores in 2006 are slightly larger than those in 2002.

Table 4.5 Descriptive Statistics of Civic Engagement Scores in Four CIRP Administrations

	N	Mean	S.D.	Min.	Max.
TFS02	355,503	44.50	14.82	-6.71	100.60
TFS06	331,158	46.48	15.58	-6.81	100.59
CSS02	34,494	47.73	15.05	-7.75	101.10
CSS06	28,077	49.64	15.50	-7.79	100.87

The distributions of these civic engagement scores in each CIRP administration are also shown in box-and-whisker plots. In Figure 4.2, the band near the middle of the box is the median. The bottom and the top of the dark middle boxes represent the lower and upper quartiles (i.e., 25th percentile and 75th percentile). The T-bars (or whiskers) extend to 1.5 times the height of the box, and the circles and stars falling outside of the whiskers are outliers. The stars are extreme outliers because those cases have values more

than three times the height of the boxes. For the ease of comparison, the long thin line in the middle of the graph indicates the overall mean of civic engagement scores (45.72). Figure 4.2 shows that the medians of civic engagement scores are also higher among senior students than those among freshman students. Some outliers are identified, but they represent a small percentage of the entire student sample.

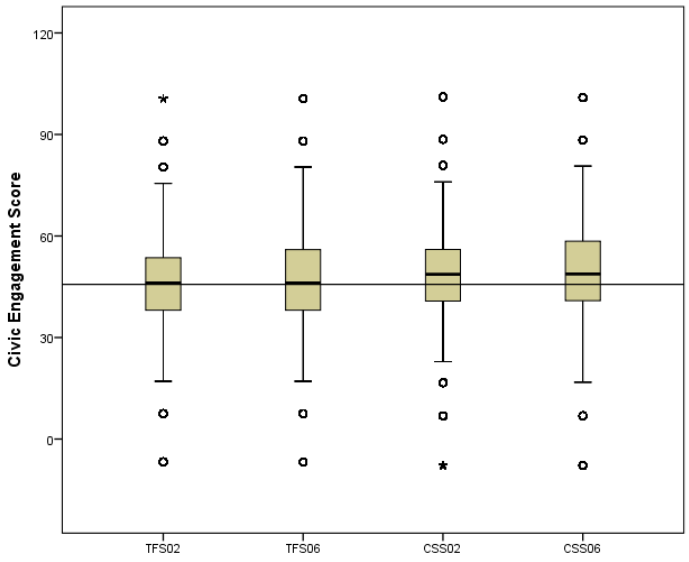


Figure 4.2 Boxplots of Civic Engagement Scores in Four CIRP Administrations

For the subsample of students who participated in both TFS02 and CSS06 surveys, their civic engagement mean scores increased from 45.04 (S.D. = 13.35) in the freshman year to 49.07 (S.D. = 14.98) in the senior year. The average change in civic engagement scores is 4.03 (S.D. = 14.80). Figure 4.3 presents two ways of visually examining this change. In the plot on the left, civic engagement scores in the freshman year are represented by light bars. Civic engagement scores in the senior year are represented by

dark bars. The increase of the civic scores is shown from the dark bars shifting to the right of the light bars, and the peak of the light bars indicates a lower civic engagement score than that of the dark bars. The plot on the right represents the distribution of the same student's changes in civic scores from 2002 to 2006. The peak of the graph is slightly below zero (around -0.06), but 60% of the students scored higher in civic engagement at the end of their four years of college study. There is a moderate correlation between students' civic engagement scores in the freshman year and the senior year ($r = 0.46$, $p < .001$). This shows that students' responded consistently to the same 8-item scale, but their responses are far from identical. After the four college years, most students put a higher value on (and developed positive attitudes towards) civic engagement.

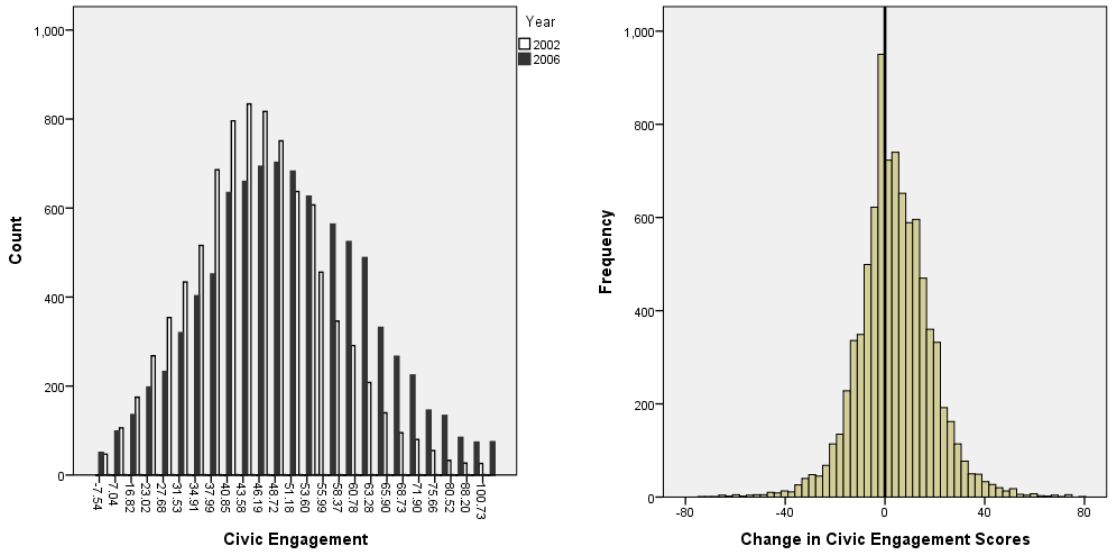


Figure 4.3 The Longitudinal Changes of the Civic Engagement Scores

4.1.2 Student-level Variables

As described in chapter three, four types of student-level variables were included in this study. Namely, student and family background characteristics, high school educational outcomes and educational aspirations, college educational outcomes and experiences, and student profile. Most of the background characteristics and educational outcomes are straightforward. This section focuses on the construction of the student profile variable and a post-stratification weight variable.

First, latent class analysis was used to explore the optimal number of classes of student characteristics among freshman students in the longitudinal dataset. During the analysis, multiple models were conducted to test increasing numbers of latent classes. From the CIRP freshman survey in 2002, 34 items regarding students' activities, perceived goals in life, and self-ratings upon entry to college were used in the models. To avoid potential multicollinearity in future regression analyses, civic engagement related items were not selected. LCA fit indices for two to eight class solutions are summarized in Table 4.6. As the table shows, the drop in G^2 relative to the drop in degrees of freedom was faster when the number of classes was smaller than five. When the number of classes was greater than five, there was still some improvement in model fit, but the drop in G^2 comparative to degrees of freedom was much slower. The AIC and BIC values also agreed with the G^2 statistic. Additionally, the entropy R^2 index indicates how well a model classifies the cases in the data, with a value close to 1 indicating a good classification. In this case, the five-class classification was shown to provide better fit among all the models tested.

Table 4.6 Model Fit Indices for 2- to 8- Class Solutions for the Student Profile Variable

No. of Classes	Likelihood Ratio G^2	Degrees of Freedom	AIC	BIC	Entropy R^2
2	366867	7699	367177	368265	0.84
3	360461	7621	360927	362562	0.81
4	355908	7543	356530	358713	0.83
5	**352765**	**7465**	**353543**	**356274**	**0.84**
6	350341	7387	351275	354553	0.84
7	348318	7309	349408	353233	0.84
8	346530	7231	347776	352148	0.83

Note. The selected model is bolded. AIC = Akaike's Information Criterion; BIC = Bayesian Information Criterion.

Finally, balancing between model fit, parsimony, and interpretability, a five-class model emerged as the most plausible among all the solutions. Based on the estimated probability of membership for each case, and the item-response probabilities for endorsing each item, the five student profiles are named (1) Status Striver; (2) Artist; (3) Hedonist; (4) Religious-oriented; and (5) Disengaged.

Status strivers comprise 17% of the students. Similar to the status strivers in Astin's typology, these students show a strong drive to achieve. Compared to their peers, it is more important for them to become an authority in their field, obtain recognition from colleagues for contributions they make, and become successful in a business of their own. They already had high self-confidence in intellectual ability but still spent relatively more time studying/doing homework and talking with teachers outside of class. College is important to them as preparation for graduate or professional school. At the same time, it is also important for them to raise a family and be accomplished in performing arts such as acting and dancing (although they are behind artists with regard

to the importance of the latter). Except for artists, they have more financial concerns about college education than their peers. They spent the most time working for pay during the last year in high school and aspire to be very well off financially.

Artists comprise about 14% of the students in the longitudinal data. They are distinct from their peers in highly rating themselves on artistic ability and creativity. Creating artistic works (e.g., painting, sculpture, and decorating) and writing original works (e.g., poems, novels, and short stories) are most important. They also spent an above average amount of time playing musical instruments during the last year in high school.

Hedonists comprise about 24% of the students. They are the group of students who most frequently drank beer, wine or liquor, smoked, and partied during the last year of high school. They also have the least amount of concern for financing their college education.

The religious-oriented students comprise about 25% of the student sample. Among all students, they have the highest self-ratings on religiousness and spirituality. They also most frequently discussed religion and attended a religious service. Moreover, they spent the least amount of time partying and are the least likely to drink beer and liquor or smoke. They played musical instruments quite often during the last year of high school. Other than hedonists, they have the least amount of financial concern about college. They rated themselves highly on academic ability (only lower than status strivers), but they do not have strong desires to become an authority in their field, to

become successful in a business of their own, to obtain recognition from colleagues for their contributions to the field, or to make a theoretical contribution to science.

The last group of students is named disengaged and comprises about 20% of the sample. They do not consider themselves to be religious or spiritual. They are not interested in creating artistic works or writing original works. They also rated themselves the lowest on having intellectual self-confidence and were not overwhelmed by what they had to do.

To explore the relationship between civic engagement and the student profile variable, Table 4.7 summarizes students' unadjusted civic engagement scores at the freshman year, the senior year, and changes in civic engagement scores among these five classes. The table shows that disengaged students scored lowest on the civic engagement measure among all students in both the freshman and senior years, but they had the second largest average increase in civic

Table 4.7 Means and Standard Deviations (in parentheses) of Civic Engagement Scores and Changes by Student Profile

	Disengaged	Status Strivers	Artists	Hedonists	Religious-oriented
Civic Engagement in Freshman Year	39.90 (13.33)	52.86 (12.00)	48.44 (13.79)	42.37 (12.68)	44.23 (11.61)
Civic Engagement in Senior Year	44.92 (15.16)	54.02 (14.61)	51.20 (15.53)	47.19 (14.91)	49.45 (13.56)
Changes	5.02 (15.56)	1.17 (14.38)	2.76 (15.39)	4.82 (15.17)	5.22 (13.41)

engagement. Conversely, status strivers scored the highest at both grade levels and their average score increase was the smallest among the five groups.

Figure 4.4 represents box-and-whisker plots of civic engagement scores in 2002, 2006, and changes in civic engagement scores for each student profile category. Two reference lines are added. One indicates the mean civic engagement score in 2002 among all students (45.04). The other is located at zero. It is worth noting that disengaged students' median scores were the lowest in civic engagement in both 2002 and 2006. Their median score in the senior year was even below the 25th percentile score for status strivers. Moreover, many more outliers were identified in the change scores than in the freshman and senior civic engagement status scores.

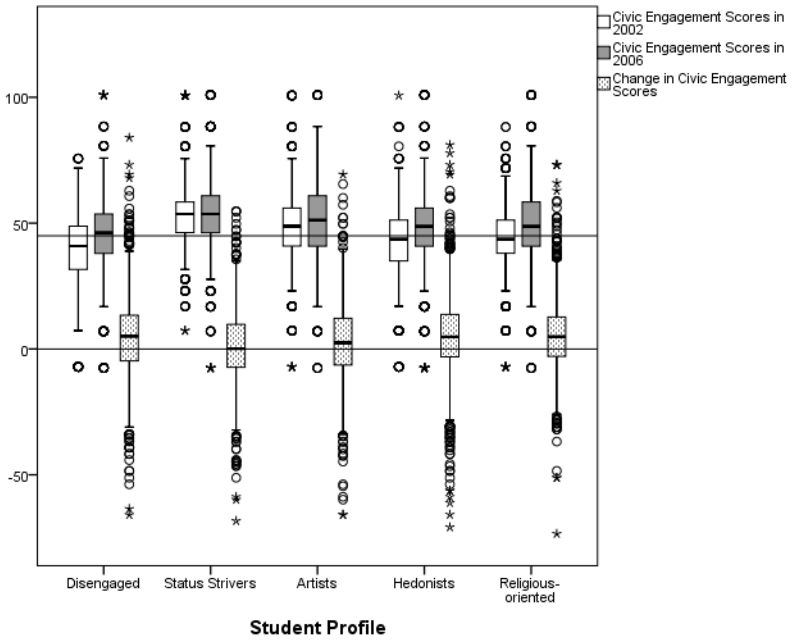


Figure 4.4 Boxplots of Civic Engagement Scores and Score Changes by Student Profiles

Because students usually participated in CIRP surveys on a voluntary basis, post-stratification is considered in this study to reduce biased estimates due to nonrepresentativeness in the voluntary survey sample. IPEDS collects information on each institution's total student enrollment, as well as the proportions of students in each of the ten gender by race/ethnicity subgroups (two subgroups for gender, male and female; five subgroups for race/ethnicity, White, African American, Asian American, Hispanic, and others). Differences in the percentages of specific demographic subgroups in the population and the sample in the final analysis were compared and presented in Table 4.8. Significant differences were found in both gender and race/ethnicity subgroups. In general, the final sample included more female students. Among race/ethnicity subgroups, white students were overrepresented, while African American, Asian

Table 4.8 Percentages of Selected Demographic Subgroups in the Institutional Population and the Sample

Demographic Subgroup	% in Population	% in Sample	*p*-value
Gender			
Male	43.4	35.2	< .001
Female	56.6	64.8	
Race/Ethnicity			
White	74.0	82.1	< .001
African American	5.2	2.0	
Asian American	7.2	5.3	
Hispanic	6.5	2.9	
Others	7.1	7.7	

American and Hispanic students were underrepresented. Therefore, post-stratification weights were created based on gender, race/ethnicity, and institution strata. These post-stratification weights were used throughout the analyses in modeling the associations between student- and institution-level covariates and students' civic engagement scores.

4.1.3 School-level Variables

Many school level variables are available in the dataset including Carnegie classification variables from the IPEDS, and other institutional characteristics collected by HERI such as institution type, institution control, and selectivity. Based on selected variables, Table 4.9 lists civic engagement mean scores and score changes at the institutional level among different institutional classifications. In order to compare the same students' changes in civic engagement, only students who were matched in the longitudinal dataset were included in the table.

One thing worth noting in Table 4.9 is that the means of the civic engagement scores were always higher in the senior year than those in the freshman year, regardless of institution type. Comparing across the three combined categories in the basic Carnegie classification[19], the mean of the civic engagement scores in the freshman year in baccalaureate colleges was slightly higher than those in master's and doctoral/research institutions. However, baccalaureate colleges had the smallest changes in civic

[19] *Baccalaureate Colleges* include "institutions where baccalaureate degrees represent at least 10 percent of all undergraduate degrees and where fewer than 50 master's degrees or 20 doctoral degrees were awarded during the update year." *Master's Colleges and Universities* include "institutions that award at least 50 master's degrees and fewer than 20 doctoral degrees during the update year." *Doctorate-granting Universities* include "institutions that awarded at least 20 research doctoral degrees during the update year," but exclude "doctoral-level degrees that qualify recipients for entry into professional practice," such as the JD and MD (Carnegie Foundation for the Advancement of Teaching, n.d.a).

Table 4.9 Civic Engagement Scores and Changes (Mean and Standard Deviation) among Different Types of Institutions

	Number of Institutions	Freshman Civic Engagement	Senior Civic Engagement	Changes
By Basic Carnegie Classification				
Baccalaureate Colleges	43	45.45 (3.78)	49.24 (3.96)	3.79 (3.51)
Master's Colleges and Universities	38	44.58 (3.36)	49.31 (2.42)	4.73 (2.84)
Doctoral/Research Universities	9	44.56 (2.85)	49.10 (3.53)	4.55 (2.16)
By Institutional Control and Religious Type				
Public	14	44.53 (2.97)	49.29 (3.06)	4.76 (2.18)
Private/Nonsectarian	28	44.69 (3.41)	48.37 (3.37)	3.68 (3.36)
Catholic	13	45.79 (2.76)	49.79 (2.31)	3.99 (2.24)
Other Religion	35	45.13 (4.08)	49.76 (3.62)	4.63 (3.54)
By Institutional Selectivity				
Low	4	43.92 (1.66)	48.48 (1.44)	4.56 (1.13)
Medium	33	44.88 (3.43)	48.60 (3.62)	3.72 (3.20)
High	18	45.18 (2.91)	49.65 (2.09)	4.47 (2.42)
Very High	35	45.13 (4.08)	49.76 (3.62)	4.63 (3.54)
By Carnegie Classification: Undergraduate Instructional Program				
Arts and Sciences Focus	16	47.27 (2.33)	49.73 (2.69)	2.46 (3.28)
Arts and Sciences plus Professions	17	44.91 (3.33)	50.80 (4.20)	5.89 (2.86)
Balanced Arts and Sciences and Professions	32	45.44 (3.88)	49.74 (2.57)	4.30 (3.21)
Professions plus Arts and Sciences	23	43.31 (2.68)	47.70 (2.41)	4.38 (2.39)

engagement compared with master's and doctoral/research institutions. Similarly, on average, Private/Nonsectarian institutions had the smallest changes in civic engagement compared with public institutions and institutions that have a religious affiliation. The

changes in civic engagement were similar across institutions when they were classified by selectivity level.

Results from ANOVA (not shown in the table) further demonstrated that the differences in civic engagement scores and score changes were not statistically significant ($p > .05$) for most types of classifications. However, statistically significant differences were found in civic engagement scores and score changes when institutions were classified by their undergraduate instructional programs[20] ($Civic_F$: $F_{(3,84)}$=4.86, p<.01; $Civic_S$: $F_{(3,84)}$=4.10, p<.01; $Civic_{Change}$: $F_{(3,84)}$=3.70, p=.02). Specifically, Tukey's post-hoc tests showed that institutions with a higher percent of bachelor's degree majors in the Arts and Sciences (i.e., at least 80 percent) had a significantly smaller increase on civic engagement compared with institutions with a lower percent of bachelor's degree majors in the Arts and Sciences (i.e., 60-79 percent) but a larger percent in professional fields, such as business, education, engineering, health, and social work.

Since institutions were compared within more homogeneous peer groups in this study, these institutional level variables were also used to categorize institutions. Based on variables such as institution size (i.e., number of full-time equivalent and degree-seeking students enrolled in the institution), institutional type and selectivity level, basic Carnegie classification, undergraduate instructional programs of the institution, and institutional enrollment profile, institutions were categorized into several peer groups.

---

[20] According to the degree data, *Arts & Sciences Focus* institutions have "at least 80 percent of bachelor's degree majors" in the Arts and Sciences; *Arts & Sciences plus Professions* institutions have "60-79 percent of bachelor's degree majors" in the Arts and Sciences; in *Balanced Arts & Sciences/Professions* institutions, "bachelor's degree majors were relatively balanced between arts and sciences and professional fields (41-59 percent in each)"; in *Professions plus Arts & Sciences* institutions, "60-79 percent of bachelor's degree majors were in professional fields (such as business, education, engineering, health, and social work)" (Carnegie Foundation for the Advancement of Teaching, n.d.b).

For example, nine peer groups were generated among institutions that participated in all four CIRP survey administrations. To ensure institutional anonymity, each of the peer groups has at least five institutions. Table 4.10 describes these peer groups.

Table 4.10 Peer Groups among Institutions that Participated in All Four CIRP Surveys

	Number of Institutions	Description[a]
Peer group 1	11	Small 4 year baccalaureate institutions; Arts & Sciences focus undergraduate instructional programs; private, high selectivity level
Peer group 2	7	Small 4 year baccalaureate institutions; Arts & Sciences plus professions undergraduate instructional programs; private, high selectivity level
Peer group 3	7	Small 4 year baccalaureate institutions; balanced Arts & Sciences and professions undergraduate instructional programs; other religious affiliations (not Catholic), mid to high selectivity level
Peer group 4	8	Small 4 year baccalaureate institutions; professions plus Arts & Sciences undergraduate instructional programs; other religious affiliations (not Catholic); located in the great lakes area
Peer group 5	10	Medium 4 year Master's institutions; balanced Arts & Sciences and professions undergraduate instructional programs; private
Peer group 6	9	4 year Master's institutions; professions plus Arts & Sciences undergraduate instructional programs
Peer group 7	5	Private, 4 year doctorate/research institutions
Peer group 8	8	Private, 4 year institutions; high to very high selectivity level; located in the far west area
Peer group 9	7	Private, 4 year institutions; low selectivity level; located in the Mid-Atlantic area

[a] Based on Carnegie classification, an institution is classified as small when it has 1,999 or less FTE and degree-seeking students enrolled, a medium-sized institution has 2,000-4,999 such students enrolled, and a large-sized institution has 5,000 or more such students enrolled.
Institutional type and selectivity levels are based on HERI's classification named stratification cell. More details of this classification are available at http://www.heri.ucla.edu/researchers/Strat.Cell.Comp.Group.PDF.

4.2 Research Question One

To explore the associations between undergraduate students' characteristics and their civic engagement scores in the senior year after adjusting for scores in their freshman year, the longitudinal dataset with matched students' responses from both the freshman year and the senior year was used. Hierarchical linear models were carried out in four stages. In the first stage, an unconditional model with no predictors at either level of the hierarchy was built. The intraclass correlation coefficient (ICC) is 0.03 (i.e., $\frac{\tau_{00}}{\sigma^2 + \tau_{00}} = \frac{6.75}{220.25 + 6.75}$), which means that only three percent of the total variance in senior students' civic engagement scores as measured in the 2006 CIRP senior survey was between institutions. Given that NSSE has reported similarly small between-institution variances (4%-8%, with one exception of 14%) in all five NSSE benchmarks on student engagement, this is not a surprising finding (NSSE, 2008). Although students are more variable within institutions than between institutions, multilevel models were used throughout the study because of their ability to account for the nesting structure of the data where students were nested in institutions, as well as the ability to provide appropriate standard error estimates and the statistical significance level of the results. Moreover, a simulation study has found that the assumption of independent observations can still be violated even when the ICC is small (Roberts, 2007). Comparing the use of a hierarchical linear model and an OLS model, the likelihood ratio test also found significant improvement in the model fit in the hierarchical linear model (χ^2=147.15,

p<.001). The statistically significant χ^2 suggests that the grouping structure is relevant for the analysis.

At stage two, the civic engagement scores of the same students from their freshman year were group-mean centered and added to the model after adjusting for measurement error (see section 3.3.2). The aggregated mean of freshman civic engagement scores for each institution was also added at level-2 so that differences in institution mean scores were controlled for in the model. As expected, results showed a strong positive association between students' civic engagement scores upon entry to college and four years later (0.63, $t(8782) = 38.00$, $p < .001$).

At stage three, student-level covariates were first added by group with the use of a less stringent significance level (i.e., .05). After examining all the variables by group, all retained variables were added to the next model and only those significant at the .01 level were kept in this model. If a random level-1 slope could be reliability estimated, it was allowed to vary across institutions. Results from the final model at this stage are presented under Model 3 in Table 4.11.

The fixed regression coefficients are at the top of the table. After adjusting for all other student covariates, students' civic engagement scores upon entry to college and four years later were still strongly associated (0.52, $t(8746) = 31.38$, $p < .001$). Among the first group of student covariates, race/ethnicity subgroups were the only statistically significant ones. Other student demographics and family background covariates, such as gender, native language status, religion, and an indicator of socioeconomic status, were not significantly associated with students' civic engagement measures in the senior year

after controlling for those in their freshman year. Among race/ethnicity subgroups, the mean difference between Hispanic students and white students was 2.47 points. This means that in a typical institution, Hispanic students scored higher in the civic engagement measure in their senior year than their white counterparts after adjusting for their freshman civic measures, academic achievement levels, and other student characteristics (2.47, $t(8746)=3.06$, $p=.003$). Similarly, students who are multiracial and whose race/ethnicity are unknown or fall in other categories (i.e., the race/ethnicity subgroup "others") also scored 1.86 points higher than their white counterparts (1.86, $t(8746)=3.34$, $p=.001$).

The second group of student-level covariates entered into the model includes students' educational aspirations, their SAT/ACT scores, and high school GPAs. All these covariates were statistically significant in the final model. This means in the average postsecondary institution, compared with students who intended to obtain a bachelor's degree as their highest academic degree, those who intended to earn a degree that requires more advanced studies scored higher in their senior civic engagement level after controlling for all other covariates in the model. Specifically, those who intended to earn a law degree scored the highest (4.48, $t(6315)=5.99$, $p<.001$), followed by those who intended to earn a degree in divinity (4.15 $t(8746)=3.12$, $p=.002$). Given that the standard deviation of the civic engagement measure is less than 15, being a student who intended to obtain a professional degree in law or divinity rather than those who only intended to obtain a bachelor's degree, the student's score in civic engagement increased by nearly

Table 4.11 Results of HLMs for Research Question One

	Model 3			Replicated Model 3		
Level-1 Fixed Effects[a]	Coef.	s.e.	*p*-value	Coef.	s.e.	*p*-value
Intercept, γ_{00}	49.31	0.21	**<.001**	48.62	0.44	**<.001**
Civic Engagement in the Freshman Year, γ_{10}	0.52	0.02	**<.001**	0.52	0.02	**<.001**
Student Demographics & Family Background						
Race/Ethnicity[b]						
African American, γ_{20}	0.91	1.15	.431	1.02	1.16	.378
Asian American, γ_{30}	1.35	0.68	.046	1.35	0.74	.067
Hispanic, γ_{40}	2.47	0.81	**.003**	2.42	0.80	**.003**
Others, γ_{50}	1.86	0.56	**.001**	1.55	0.66	.020
High School Educational Outcome & Aspiration						
SAT/ACT, γ_{60}	-0.01	0.00	**<.001**	-0.01	0.00	**<.001**
HSGPA, γ_{70}	-1.70	0.46	**<.001**	-1.58	0.57	**.006**
Educational Aspiration[c]						
Master's Degree, γ_{80}	1.07	0.38	**.005**	0.94	0.54	.081
Law Degree, γ_{90}	4.48	0.75	**<.001**	4.07	0.79	**<.001**
Medicine Degree, γ_{100}	1.73	0.78	.026	1.14	0.84	.174
Divinity Degree, γ_{110}	4.15	1.33	**.002**	2.98	2.55	.243
Ph.D./Ed.D., γ_{120}	2.42	0.53	**<.001**	2.04	0.66	**.002**
Others, γ_{130}	0.26	1.08	.807	1.03	1.09	.345
College Educational Outcomes and Experiences						
College GPA, γ_{140}	1.43	0.48	**.004**	1.63	0.64	.018
Participation in civic related activities in senior year, γ_{150}	2.47	0.13	**<.001**	2.39	0.14	**<.001**
Student Profile[d]						
Religious Oriented, γ_{160}	1.90	0.46	**<.001**	1.74	0.60	**.004**
Status Strivers, γ_{170}	2.09	0.54	**<.001**	2.81	0.59	**<.001**
Artists, γ_{180}	2.30	0.58	**<.001**	3.03	0.61	**<.001**
Hedonists, γ_{190}	1.39	0.46	**.003**	1.62	0.53	**.003**
Random Effects	Variance Component	df	χ^2 *p*-value	Variance Component	df	χ^2 *p*-value
Mean Civ. Eng., u_{0j}	2.63	70	182.51 **<.001**	3.97	6	50.24 **<.001**
College GPA slope, u_{14j}	4.03	89	122.87 **.010**	3.78	25	46.21 **.006**
Level-1 effect, r_{ij}	161.58			167.70		
Reliability Estimate						
Intercept	0.517			0.810		
College GPA	0.257			0.447		

Note. Bolded values are significant at .01

[a] Fixed effects at level 2 is not reported here, since they are not of research interests in research question one; [b] Reference group is White; [c] Reference group is Bachelor degree; [d] Reference group is Disengaged

one-third of the standard deviation of the civic engagement measure after four years of college experience.

As for students' educational outcomes in high school, both SAT/ACT scores and high school GPAs were negatively associated with senior students' civic engagement scores (SAT/ACT: -0.01, $t(8746)=-4.83$, $p<.001$; HSGPA: -1.70, $t(8746)=-3.69$, $p<.001$). In other words, the higher a student's SAT/ACT score and high school GPA were, the lower the student's civic engagement level was at his/her senior year. Because of the use of group-mean centering, the beta coefficient is interpreted somewhat differently from that in grand-mean centering. Take SAT/ACT as an example, the beta coefficient associated with SAT/ACT ($\hat{\gamma}_{60}$ = -0.01) indicates that, for every 100 point (or 1 standard deviation) increase in a student's SAT/ACT score above and beyond the mean SAT/ACT score of the school the students attended (rather than the grand mean of the SAT/ACT score across all students and schools), the student's civic engagement score decreased by 1 point (i.e., -0.01×100) at the end of the senior year, given all other student background controlled for are similar. This statistical significance is probably due to the large sample size in the analysis. Yet, the change is too small to have practical significance.

It is interesting to note that while students' SAT/ACT scores and high school GPAs are negatively associated with the civic engagement measures, their GPAs in college are positively associated with the changes (1.43, $t(89)=3.00$, $p=.004$). For every one point increase in college GPA (i.e., one letter change such as from C to B or from B to A) above and beyond the school mean, students' civic engagement measures increased by 1.43 points. However, it is not clear why civic engagement scores are negatively

associated with academic achievement measures from before entering college but positively associated with academic achievement measures during college. But again, these changes in students' civic engagement scores are too small and are not of substantial importance.

The other variable in the third group of student-level covariates is an indicator of students' civic related activities during college. It is the sum of students' responses on three items – time spent on student clubs/groups, how often did you perform volunteer work, and how often did you vote in a student election. All these items were on a 3-point scale with the higher points identifying more frequent participation in those civic-related activities. This civic participation indicator was positively associated with students' attitudes and beliefs in civic engagement as well (2.47, $t(8746)=18.80$, $p<.001$); so, the more frequent participation in these activities was, the more positive a student's attitudes and beliefs towards civic engagement in his/her senior year was.

Student profile is the last student-level covariate that entered into the model. The results show that, controlling for all other covariates, students with different profile categories perform significantly differently in civic engagement. Compared to their peers, disengaged students scored significantly lower in civic engagement in their senior year. Among the other four student types, the regression coefficient for artists is the highest (2.30, $t(8746)=3.99$, $p<.001$), followed with status strivers (2.09, $t(8746)=3.87$, $p<.001$), religious oriented students (1.90, $t(8746)=4.11$, $p<.001$), and hedonists (1.39, $t(8746)=3.01$, $p=.003$).

Below the level-1 fixed effects section, Table 4.11 lists the estimated variances of the random effects at both levels. The residual variance within institutions in model three, 161.58, is smaller than 220.25 in the unconditional model. The proportion of variance accounted for within institutions is

$$\frac{\hat{\sigma}^2(unconditional) - \hat{\sigma}^2(conditional)}{\hat{\sigma}^2(unconditional)} = \frac{220.25 - 161.58}{220.25} = 0.27$$

In other words, 27% of the variance within institutions was explained in model 3. Table 4.11 also lists the variance components in mean civic engagement scores among seniors (i.e., 2.63) and in the mean college GPA – civic engagement slope (i.e., 4.03). The significant χ^2 statistic associated with mean civic engagement (p<.01) indicates that significant differences existed among the institutional means on senior students' civic engagement levels. The χ^2 statistic associated with the mean slope was also significant (p=.01). This indicates that the relationship between college GPA and civic engagement measures in the senior year within institutions varied significantly across all institutions. Also of note is that this random slope had a reliability coefficient of 0.257. Therefore, the college GPA – civic engagement slope was allowed to vary across institutions in later phases of the study.

Finally, interaction effects between student-level covariates were explored in stage four. A series of interaction terms were created between the covariates identified as statistically significant in stage three and other covariates. Namely, the interaction terms included five racial/ethnic subgroups and five student profile types by freshman civic engagement scores, by SAT/ACT scores, by high school GPA, by college GPA, and by

civic related activities during college, as well as interactions between freshman civic engagement scores, SAT/ACT scores, high school GPA, college GPA, and civic related activities. However, none of the interaction terms was found to be statistically significant. Therefore, model three is the final model used to answer research question one.

Table 4.12 Variance Decompositions for Senior Students' Civic Engagement

			Within institutions		Between institutions	
Model	Level 1 covariates	Level 2 covariates	Variance	Percent of variance in model 1 accounted for	Variance	Percent of variance in model 1 accounted for
1	None	None	220.25	-	6.75	-
2	Freshmen's civic engagement	None*	176.19	20	3.04	55
3	Freshmen's civic engagement and other student characteristics	None*	161.58	27	2.63	61

* Since level 1 covariates were group-mean centered to obtain accurate student-level estimates, the institutional means on these covariates were added at level 2 in the statistical model (to acquire proper variance decompositions). But they were not listed in the table to avoid potential confusion.

Table 4.12 summarizes the variance decompositions in the final model at each stage and the percent of variance in the baseline model accounted for both within and between institutions. Model 1 is the unconditional model or the baseline model. As discussed earlier, only 3% of total variance in the civic engagement scores among senior students was between institutions. Most of the variation existed among students, within institutions. In model 2, because of the strong association between students' civic engagement scores in the freshman year and senior year (r = .46), the addition of students'

civic engagement scores in the freshman year accounted for 20% of the variance within institutions and 55% of the variance between institutions. In other words, 21% of the total variance in senior students' civic engagement scores (i.e.,

$$\frac{Total\,Variance - Residual\,Variance}{Total\,Variance} = \frac{(220.25 + 6.75) - (176.19 + 3.04)}{220.25 + 6.75} = 0.21)$$ was

accounted for by their prior civic engagement levels. In model 3, all the student covariates as listed in Table 4.11 were included in the model. This model explained 27% of the total variance within institutions and 61% of the total variance between institutions. Comparing with model 2, an extra 7 percentage point increase in the variance within institutions (i.e., 27% - 20%) and 6 percentage point increase in the variance between institutions (i.e., 61% - 55%) was accounted for. That is 28% of the total variance in senior students' civic engagement scores.

Because only students who were in the same postsecondary institutions in their freshman and senior years and who also had matching records in both the CIRP freshman survey in 2002 and the senior survey in 2006 were included in the longitudinal dataset, the sample sizes within each institution could be small. In the above analysis, all institutions had at least 20 students in order to use the most available cases. To obtain more reliable estimates, the same multilevel model was replicated for institutions that had at least 100 matched students in the dataset. The number of institutions changed from 90 to 26 in the replicated analysis. The results are reported under *Replicated Model 3* in Table 4.11.

As expected, the reliability estimates were higher in the replicated model with a minimum of 100 students per institution. For the estimation of the intercepts, the reliability coefficients increased from 0.52 to 0.81. For the estimation of the slope associated with college GPA, the reliability coefficients increased from 0.26 to 0.45. Table 4.11 also confirmed that the estimated regression coefficients in both models were similar. Most covariates remained statistically significant at the .01 level in both models. Small differences existed in estimates of college GPA and the "others" racial/ethnic category. Specifically, the *p*-values for college GPA increased to .018. Similarly, the differences in senior students' civic engagement between white students and students in the "others" category changed from significant at the .01 level to significant at the .05 level. The biggest difference was in the estimations on the differences between planning to have a master's or divinity degree and a bachelor's degree. The estimated regression coefficients were not significant anymore ($p>.05$). For the regression coefficient associated with a divinity degree, this change was probably due to the small sample size in the divinity degree subgroup. But it was not clear on the non-significant regression coefficient associated with the master's degree subgroup.

In general, through the examinations in this section, several student-level variables were found to be associated with senior students' civic engagement scores after adjusting for their civic engagement scores as freshmen. They are students' race/ethnicity, educational aspirations, both high school and college GPAs, SAT/ACT scores, participation in civic activities in the senior year, and the student profile. The final model developed in this section can be expressed as

Level 1 (Student level):

$$Civic_{ij,S} = \beta_{0j} + \beta_{1j}\ Civic_{ij,F} + \beta_{2j}\ Black_{ij} + \beta_{3j}\ Asian_{ij} + \beta_{4j}\ Hispanic_{ij} +$$
$$\beta_{5j}\ Others_{ij} + \beta_{6j}\ SATACT_{ij} + \beta_{7j}\ HSGPA_{ij} + \beta_{8j}\ Masters_{ij} +$$
$$\beta_{9j}\ Law_{ij} + \beta_{10j}\ Medicine_{ij} + \beta_{11j}\ Divinity_{ij} + \beta_{12j}\ PhD_{ij} +$$
$$\beta_{13j}\ OtherAsp_{ij} + \beta_{14j}\ CollGPA_{ij} + \beta_{15j}\ Activities_{ij} + \beta_{16j} RelOri_{ij} +$$
$$\beta_{17j}\ Striver_{ij} + \beta_{18j}\ Artists_{ij} + \beta_{19j}\ Hedonists_{ij} + r_{ij}$$

Level 2 (Institution level):

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

$$\beta_{1j} = \gamma_{10}$$

…

$$\beta_{13j} = \gamma_{130}$$

$$\beta_{14j} = \gamma_{140} + u_{14j}$$

$$\beta_{15j} = \gamma_{150}$$

…

$$\beta_{19j} = \gamma_{190} \qquad (4.1)$$

This is the base model in the next section to answer research question two. Moreover, those identified variables, when available in the cross-sectional datasets, were used to develop value-added models to answer research question three.

4.3 Research Question Two

To investigate the association between institution-level variables and students' civic engagement scores in the senior year after adjusting for their scores in the freshman year, institutional characteristics were added at the second level of the multilevel model above (eq. 4.1). Since the regression coefficients at level-2 are the interest of this research question, this section uses a different centering approach. The level-1 covariates were grand-mean centered, continuous covariates at level-2 were also grand-mean centered, and categorical covariates at level-2 were entered into the model without centering. Institutional covariates were added one by one and those statistically significant were retained and entered into a comprehensive model together. Considering the number of institutions was relatively small, a less stringent significant level .05 was used to make decisions on statistical significance. The covariates that were not significant anymore were removed and the model was rerun.

Model 4 in Table 4.13 summarizes the results of the final model. Three institutional variables were associated with senior students' civic engagement scores beyond all level-1 covariates added in model 3. On average, senior students in master's colleges & universities scored significantly higher in civic engagement than those in baccalaureate colleges (1.40, $t(78)=2.57$, $p=.012$) given the same undergraduate instructional program type, office of education region, civic engagement scores in the freshman year, and other student background characteristics. Similarly, after controlling for all other covariates in the model, students in institutions whose undergraduate instructional program was categorized as Arts & Sciences plus professions scored higher

Table 4.13 Results of HLMs for Research Question Two

	Model 4		
Level-2 Fixed Effects[a]	Coef.	s.e.	*p*-value
Intercept, γ_{00}	48.87	0.71	**<.001**
Carnegie Classification: Basic[b]			
Master's Colleges & Universities, γ_{01}	1.40	0.55	**.012**
Doctoral/Research Universities, γ_{02}	0.41	0.58	.481
Carnegie Classification: Undergraduate Instructional Program[c]			
A&S focus, γ_{03}	0.46	0.73	.537
A&S plus professions, γ_{04}	1.48	0.56	**.011**
Professions plus A&S, γ_{05}	-0.65	0.51	.207
Others, γ_{06}	-2.01	1.31	.128
Office of Education Region[d]			
New England, γ_{07}	-0.83	0.96	.389
Mid-Atlantic, γ_{08}	-1.58	0.58	**.008**
Great Lakes, γ_{09}	-0.43	0.75	.571
West, γ_{010}	-0.34	0.67	.610
Southeast, γ_{011}	-1.27	0.82	.124

Note. Bolded values are significant at .05
[a] Fixed effects at level 1 is not reported here, since they are not of research interest at this point;
[b] Reference group is Baccalaureate Colleges;
[c] Reference group is Balanced Arts & Sciences and professions;
[d] Reference group is Far West

than those in institutions with a balanced Arts & Sciences and professions program setting (1.48, $t(78)=2.62$, $p=.011$). Students in institutions located in the Far West scored significantly higher than those located in the Mid-Atlantic area (-1.58, $t(78)=-2.73$, $p<.01$).

In terms of variance components, the residual variance between institutions decreased from 6.75 in the unconditional model to 1.80 in model 4. That is, 73% of the between-institution variance was accounted for by model 4 (i.e., $\frac{\hat{\tau}_{00}(unconditional) - \hat{\tau}_{00}(conditional)}{\hat{\tau}_{00}(unconditional)} = \frac{6.75 - 1.80}{6.75} = 0.73$). Compared with model 3 which only included student-level covariates, the addition of institutional covariates in model 4 explained an extra 12% of the residual variance remaining between institutions. As only institutional covariates were added in model 4, the residual variance within institutions was about the same as the residual variance in model 3. Since variance between institutions is a small proportion of the total variance available in senior students' civic engagement scores, these institutional covariates only explained an extra 1% of the total residual variance in model 3.

In the next stage of the analysis, cross-level interactions (i.e., interactions between institution-level variables and student-level variables) and interactions at the institution level (i.e., interactions between two institution-level variables) were explored. Interaction terms were created and tested, but no significant interaction effect emerged. Because there were only 26 institutions that had more than 100 students matched in the dataset, the analysis in this section was not replicated. The final statistical model in this section has the same level-1 equation as the final model in the previous section (eq. 4.1). The difference is the level-2 equation for β_{0j}, which can be expressed as

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(MA\_CU)_j + \gamma_{02}(Res\_U)_j + \gamma_{03}(AS\_focus)_j + \gamma_{04}(ASPro)_j + \gamma_{05}$$
$$(ProAS)_j + \gamma_{06}(AS\_Other)_j + \gamma_{07}(NewEngland)_j + \gamma_{08}(MidAtlantic)_j + \gamma_{09}$$
$$(GreatLakes)_j + \gamma_{010}(West)_j + \gamma_{011}(SouthEast)_j + u_{0j} \quad (4.2)$$

To summarize, several types of institutional subgroups were found to differ significantly in their students' civic engagement at senior year after adjusting for students' prior civic engagement scores and other student covariates. They are institution's basic Carnegie classification (or degree-granting level), the focus of undergraduate instructional programs, and geographic region. With the addition of these institutional covariates, model 4 was able to account for 73% of the variance available between institutions, which was an extra 12% of the residual variance between institutions in model 3. However, as most variance in senior students' civic engagement levels existed among students within institutions, these institutional covariates explained 28% of the total available variance in seniors' civic engagement and only an extra 1% of the total residual variance on model 3. These identified variables were used in grouping institutional peer groups but were not included in the value-added models in subsequent analyses.

4.4 Research Question Three

The third research question investigates the similarities and differences among the results of longitudinal and cross-sectional value-added models in calculating higher education institutions' value-added scores with regard to civic engagement. Two cross-

sectional value-added methods (i.e., Method 1 and Method 2) and one longitudinal method (i.e., Method 3) were used to answer this research question. This section reports the results from these value-added methods. Since different models based on different years of data were carried out using Methods 1 and 2, within-method stability is first reported after each of the two methods. After reporting results in Method 3, the overall between-method agreement and differences in all three methods and institutional value-added estimates in each peer group are further reported.

4.4.1 Method One

Results from research question one revealed several student-level variables that are associated with students' civic engagement levels. Due to the availability of the variables in the individual cross-sectional datasets, three sets of variables were included in the model for Method 1, namely, race/ethnicity subgroups, students' educational aspirations, and the sum of students' participation in three civic related activities. Because Method 1 uses institutions as the unit of analysis, all variables in the regression models were aggregated at the institution level. Therefore, the dependent variable is the institutional mean score on civic engagement among senior students. The predictors are the proportions of four race/ethnicity subgroups (i.e., African American, Asian, Hispanic, and others) within each institution, the proportions of six educational aspiration subgroups (i.e., master's degree, law degree, degrees in medicine, degrees in divinity, Ph.D/Ed.D. degree, and others) within each institution, and the mean of the civic activity participation within each institution.

As introduced in chapter three, Method 1 estimates institutional value-added scores through OLS regressions. First, two OLS regressions were carried out separately among freshman and senior students. The differences between the expected and the observed civic engagement scores (i.e., the residuals) were then saved from each of the two regression models. After putting these residuals in standard error units, the difference between the two standardized residuals was the estimated value-added score for that individual institution. Eighty-six institutions participated in both freshman and senior surveys and in both 2002 and 2006. So the analyses were replicated twice among these institutions, once for a value-added estimate based on single year data from 2002, and a second time based on single year data from 2006.

It is worth noting that OLS models rest on some key assumptions such as linearity and homoscedasticity of the error variances. The departure from these assumptions in any of the freshman and senior regression models may affect the accuracy of the institutional value-added scores. This study tested linearity, homoscadasticity, and normality assumptions for each linear regression model before calculating the value-added scores. Generally speaking, the results showed some deviations from these assumptions and some potential outliers, but all institutions remained in the value-added analyses. More detailed results are listed in Appendix C.

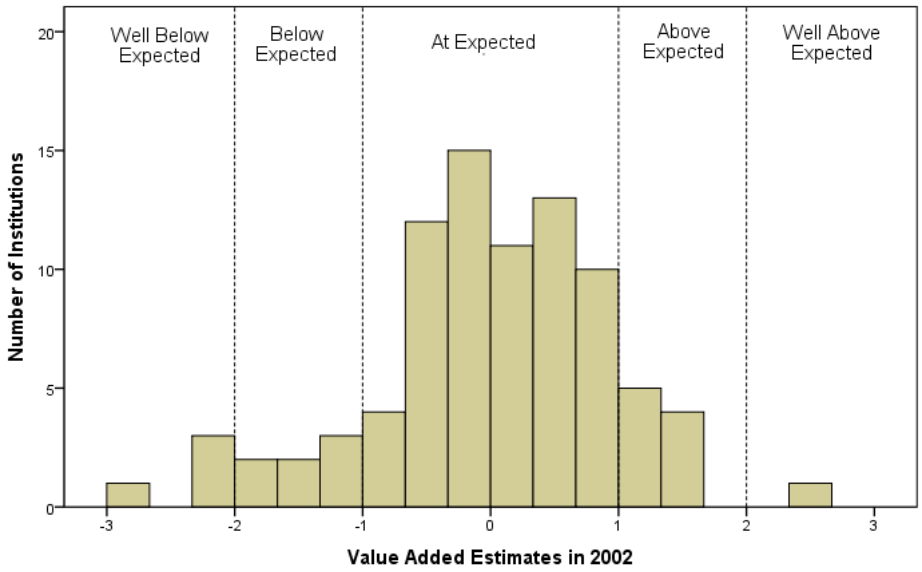


Figure 4.5 Histogram of Institutional Value-Added Estimates in 2002

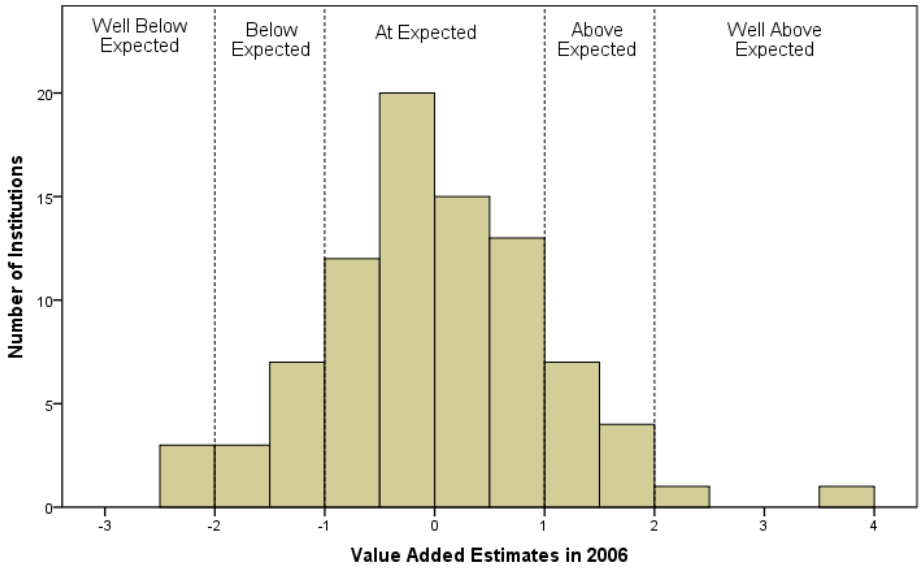


Figure 4.6 Histogram of Institutional Value-Added Estimates in 2006

The histograms of the estimated value-added scores and these institutions' performance levels are displayed in Figures 4.5 and 4.6. It shows that most institutions were at the *At Expected* level. However, several institutions were identified as *Well Below Expected* (4 institutions in 2002 and 3 in 2006) or *Well Above Expected* (1

institution in 2002 and 2 in 2006). Only one institution was identified consistently in both years as being *Well Below Expected*.

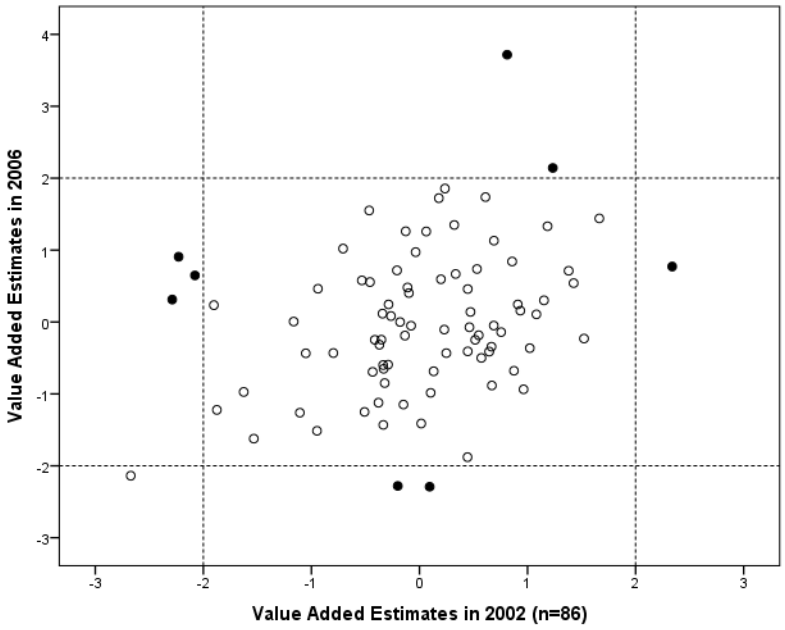


Figure 4.7 Scatterplots for Value-added Estimates in 2002 and 2006: Method One

Figure 4.7 shows the scatterplot between these two value-added estimates. The dashed lines reference the cut points between *Well Above Expected* and *Above Expected* and between *Well Below Expected* and *Below Expected* performance levels. The institution located at the lower left corner is the one that was consistently identified as *Well Below Expected* institution in both years. Using Method 1, eight of the institutions (represented by the solid dots) were identified as *Well Below (or Above) Expected* in one year, but not the other.

The Pearson correlation coefficient between the two sets of institutional value-added estimates is 0.301 (p=.005). The bootstrap is a nonparametric method for

approximating sampling variances (or standard errors), constructing confidence intervals, and conducting hypothesis tests (Efron & Tibshirani, 1993). By repeatedly sampling with replacement from the original sample, the bootstrap method obtains an empirical distribution of the statistic. Here, the bootstrap[21] method was used to estimate the correlation coefficient as well as the standard error and 95% confidence intervals of the statistic between institutional value-added in 2002 and 2006. Based on 2,000 replications, the correlation coefficient was 0.301 with a standard error of 0.09 and 95% confidence intervals of (0.113, 0.478). This coefficient remained significantly different from 0 with a *p*-value of .001.

This weak association signals far from identical institutional value-added estimates between the two years. Institutional rankings based on different value-added estimates were put into quartile subgroups and further examined in Figure 4.8. The vertical axis represents institutions' quartile groups in 2002. The horizontal axis represents these institutions' quartile groups in 2006. The percentage of institutions in the same quartile group in different value-added estimates is marked in the corresponding bar. It shows that only 38.1% of the institutions in the top quartile (i.e., 1st) in 2002 also ranked the top quartile in 2006. The percentage of institutions consistently ranked in the second, third, and forth quartiles were 13.6%, 31.8%, and 38.1%, respectively. For institutions ranked in the second quartiles in the 2002 analysis, over 86% ranked in different quartiles in the 2006 analysis. Similarly, 62% of the institutions ranked in the

[21] Jackknife is another commonly used re-sampling technique (Efron, 1982). In this case, the standard error of the correlation estimated using the jackknife method is 0.10, which is very close to the results from the bootstrap method.

top or the bottom quartiles in 2002 ranked in different quartile groups in 2006. Because institutions changed quartile groups so frequently, this can create great uncertainty, and potential injustice, if these value-added results are used in making high-stakes decision on school rewards or sanctions.

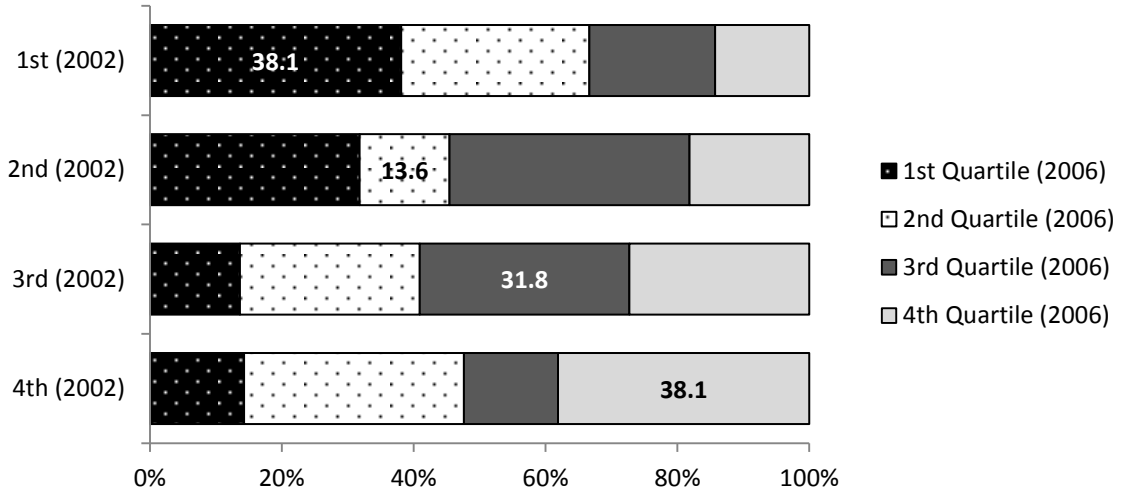


Figure 4.8 Comparisons of Institutional Quartile Group Rankings: Method One

It is hypothesized that increasing the number of respondents within institutions can increase the precision of value-added estimates. Having a representative student sample is important to make valid inferences about the participating institution as a whole. Although a relatively large sample size does not guarantee the sample to be representative, test developers usually set basic, minimum standards for student sample sizes (VSA, 2008a). For example, CAE sets it as 100 students. The previous analyses were carried out among institutions that had at least 20 students participating in each survey administration. Based on the above hypothesis, this study reexamined the three

sets of value-added estimates among 51 institutions that had at least 100 students participating in each survey administration.

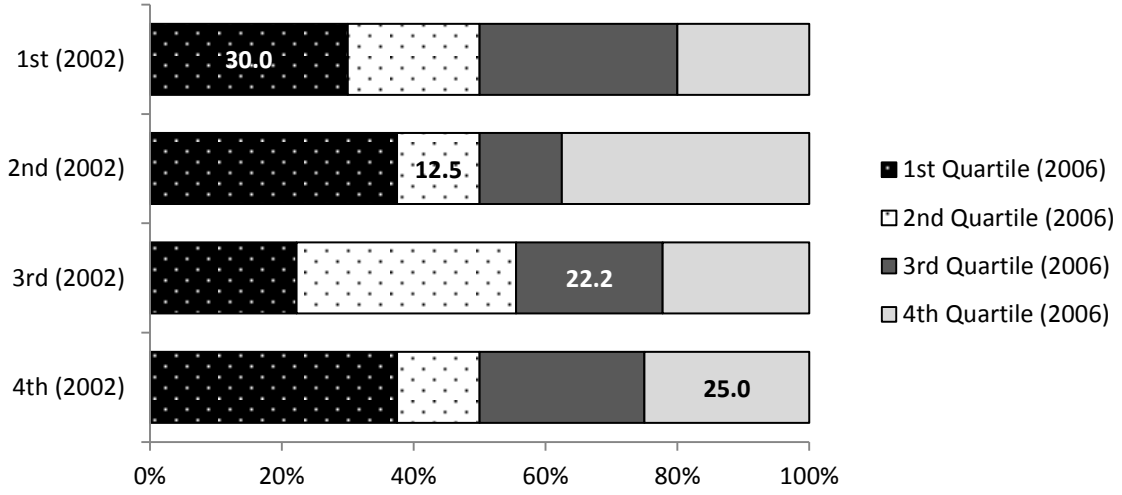


Figure 4.9 Comparison of Institutional Quartile Group Rankings between 2002 and 2006 among Institutions with 100 or More Respondents: Method One

However, results after increasing the minimum number of respondents within institutions to 100 showed a weaker correlation between two value-added estimates compared to previous analyses. The correlation between institutional value-added estimates in the two years dropped from 0.301 to 0.281 (p=.045). Using the bootstrap method based on 2,000 replications, the correlation coefficient has an increased standard error of 0.12 and 95% confidence intervals of (0.008, 0.493). Figure 4.9 compares the institutional quartile groups between the single year estimates as in Figure 4.8. The percentage of institutions consistently ranked in the same quartile groups decreased in all four quartile groups. Note that increasing the number of respondents within an institution also decreased the total number of institutions available in the comparison to only 51

institutions. This plays a role in the larger standard error of the correlation coefficient and the inconsistent rankings.

To summarize, when Method 1 was used to estimate institutional value added based on a single year of data, there was a weak year-to-year consistency. Moreover, institutional rankings based on different years of data were far from consistent. As an indicator of the within-method stability, the correlation coefficient between the two value-added estimates was only 0.30. When the minimum number of students within each institution increased from 20 to 100, the within-method stability dropped to 0.28.

4.4.2 Method Two

Multilevel models were used in value-added estimations in Method 2 to account for the dependency among students within institutions. At the student level, the same predictors in Method 1 were used, but they were not aggregated to the institution level as the unit of analysis is student. These student level variables were group mean centered at level one and institutional means on these variables were added at level two, together with the mean civic engagement scores among freshman students in the same institution. As freshman and senior students used in Method 2 are different groups of students who were studying in the same institution the same time, Method 2 is still a cross-sectional method. Therefore, the analyses were again replicated twice – once for single year estimation in 2002 and once for single year estimation in 2006. Note that HLM requires a properly specified model and makes assumptions at each level of the statistical model (Raudenbush & Bryk, 2002, chapter 9). Serious violations of these assumptions can

affect model fit and value-added estimates. The homogeneity of level-1 variances and model fit were examined and results are shown in Appendix C.

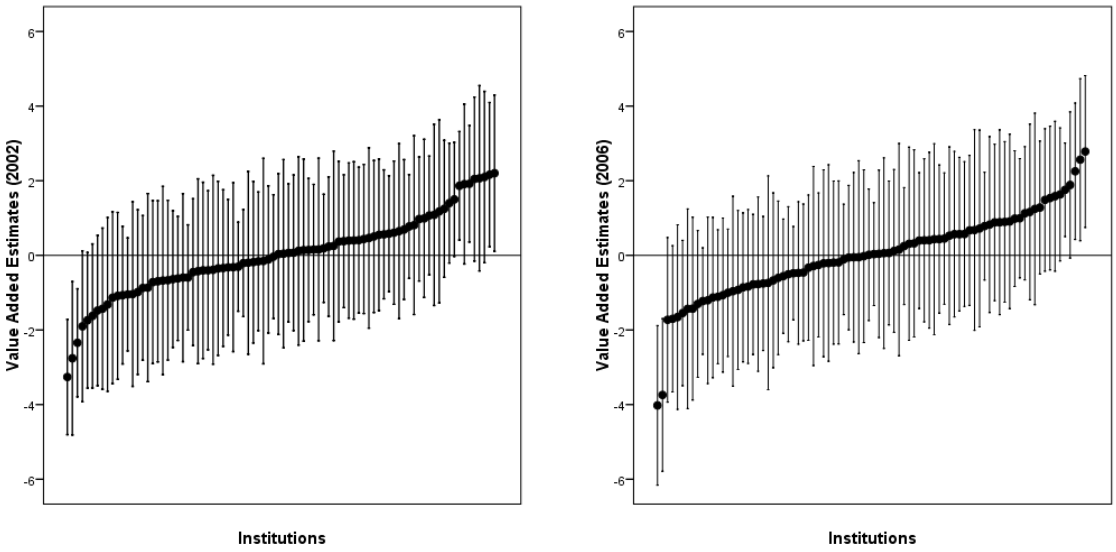
Table 4.14 Selected Results from HLMs in Method Two

	Year of 2002	Year of 2006
Total Variance Explained	11%	12%
Reliability Estimate on β_0	0.61	0.62
Value-added Estimate		
Min.	-3.26	-4.02
Max.	2.20	2.78
Std. Deviation	1.10	1.16
Range of 95% CI	4.08	4.19

Some selected results from the HLM analyses, such as the total variance explained, the reliability estimate, and descriptive statistics on the value-added estimates, are summarized in Table 4.14. Because multiple imputation was used to create five sets of data for each original dataset with missing values, each multilevel regression model was run five times. The coefficients in Table 4.14 reflect combined point and variance estimates across five multiply imputed datasets.[22]

[22] The point estimate for a parameter Q is calculated as $\overline{Q} = \frac{1}{m}\sum_{i=1}^{m}\hat{Q}_i$, where $\hat{Q}_i$ is the point estimates from the i^{th} imputed data set, i=1, 2, …, m. The variance estimate for Q is the sum of within-imputation variance and between imputation variance: $T = \overline{U} + (1+\frac{1}{m})B$, where $\overline{U}$ is the within-imputation variance, $\overline{U} = \frac{1}{m}\sum_{i=1}^{m}\hat{U}_i$; and B is the between-imputation variance, $B = \frac{1}{m-1}\sum_{i=1}^{m}(\hat{Q}_i - \overline{Q})^2$ (Rubin, 1987).

The table shows that the two models explained about the same amount of the total variance (i.e., 11% and 12%) in senior students' civic engagement scores. The reliability estimates on the level-1 intercept and the value-added estimates are also similar. The 2002 model has a slightly smaller range for the 95% confidence interval on the value-added estimate.



(a) (b)

Figure 4.10 Value-Added Estimates with 95% Confidence Intervals in Method Two

The institutional value-added estimates and standard errors associated with the value-added estimates can be further observed in Figure 4.10. These figures depict the value-added estimates based on data in 2002 (Figure 4.10a) and data in 2006 (Figure 4.10b) for each institution. The solid dots are the value-added estimates which are ordered from lowest to highest among all institutions in the dataset. The vertical bars represent the 95% confidence intervals associated with each estimate. When the

confidence interval for one institution did not include zero, it means that institution performed significantly differently from an average institution in fostering student development in civic engagement. In the analyses, 7 institutions (8%) in 2002 and 6 institutions (7%) in 2006 were identified as significantly different from the average across 86 institutions. One of those institutions was consistently identified as significantly below the average, and one was consistently above the average.

The problem of multiplicity should be noted here. Applying Tukey's philosophy on simultaneous inference (Tukey, 1953 as cited in Benjamini & Braun, 2002) to this case, when the type I error rate is set at the 0.05 level, carrying out 86 independent tests of significance using a single data set will result on average in 4.3 false positives when the null hypothesis is true. Although a slightly larger number of institutions (7 and 6) were identified in the above analyses as significantly different from an average institution, these results should be interpreted with caution.

The number of respondents within institutions was found to play a role in the value-added estimates in Method 2. Figure 4.11 is a scatterplot between the standard error of institutional value-added estimates and number of senior respondents within the institution in 2002. When the number of respondents within an institution was more than 100, the standard error of the estimates decreased to generally below 1.10.

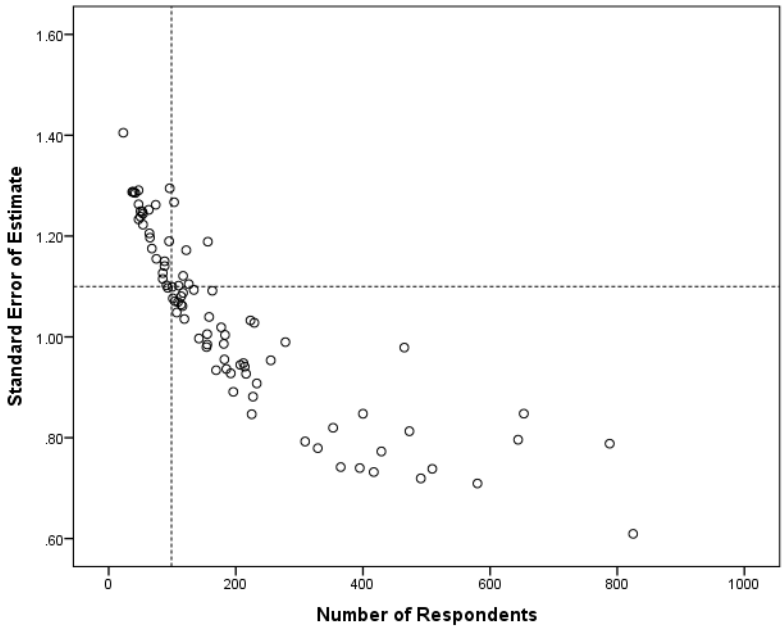


Figure 4.11 Scatterplot for the Standard Error of Value-added Estimates and Number of Respondents in 2002

Next, within-method stability was examined among analyses based on different years of data. The Pearson correlation between the two sets of value-added estimates is 0.496 (p<.01). This correlation was still not high, but it is larger than 0.301, the correlation coefficient between value-added estimates in Method 1 for the same 86 institutions. Using the bootstrap re-sampling method with 2,000 replications, the standard error of the correlation coefficient is 0.10 with a 95% confidence intervals of (0.27, 0.68). A scatterplot between the two value-added estimates is shown in Figure 4.12. The dashed line indicates the overall association between institutional value-added in 2002 and in 2006. The black dots represent institutions with at least 100 students participating in the CIRP surveys and the solid line indicates the association between the

two sets of value-added estimates among these institutions. Although the institutions with at least 100 students participating are a subset of the overall institution sample, the association between the two sets of estimates increased.

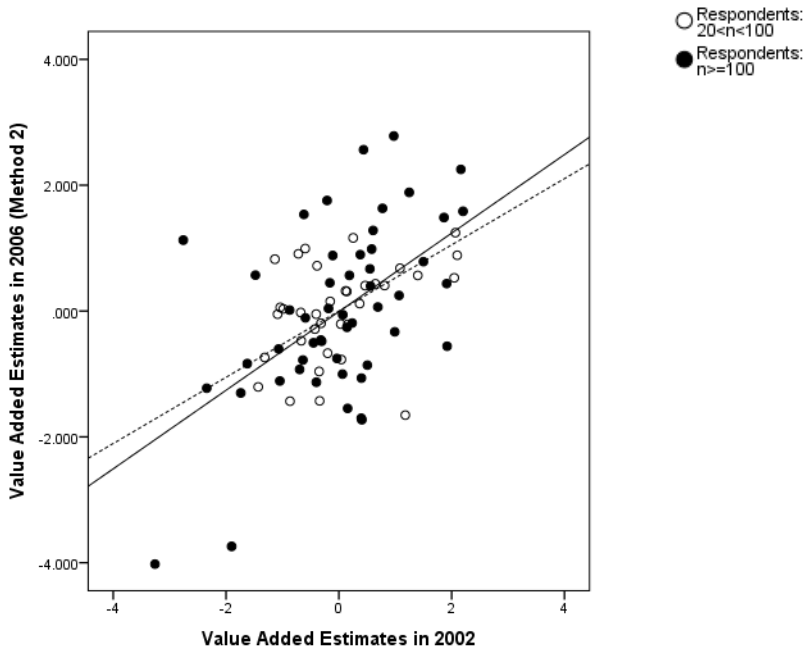


Figure 4.12 Scatter Plots for within Model Comparison in Method Two

In this section, the HLM-based Method 2 was used in estimating institutional value-added. Results from the hierarchical linear models and descriptive statistics on the value-added estimates were first summarized for Method 2. As HLM analysis has an advantage over OLS regression in providing standard errors associated with the value-added estimates for each individual institution, these standard errors were then presented and compared between different models. Only a small proportion of institutions were identified as significantly different from the average in terms of facilitating student

development in civic engagement. The number of respondents within institutions was found to play a role in within-method stabilities. Increasing the number of minimum respondents was able to decrease the standard errors of the estimates.

In general, the within-method stability was higher in Method 2 than in Method 1, as shown by the increase in the correlations between two value-added estimates from 0.30 in Method 1 to 0.50 in Method 2. More comparisons between these two methods and the comparison between these cross-sectional methods and a longitudinal method are presented in the following section.

4.4.3 Method Three

Using the longitudinal dataset with matched records for students who responded to both the freshman survey in 2002 and the senior survey in 2006, Method 3 adopted a two-level hierarchical linear model. At the student level, the same students' civic engagement scores in their freshman year and all other predictors in Methods 1 and 2 were included. At the institutional level, three sets of institutional characteristics were found in research question two to be significantly associated with students' development in civic engagement (please refer to Table 4.13 for more details). But no institutional-level predictor was included in the value-added model in Method 3. These identified institutional level variables were used in peer group classifications and the value-added estimates for each peer group are also summarized in this section. Similar to the test of HLM assumptions in Method 2, the homogeneity of level-1 variances and model fit were examined for the HLM. Detailed results are shown in Appendix C.

First, some selected results from the HLM models are shown in Table 4.15. For the purpose of comparison, similar statistical results from Method 2 using data from 2006 are also listed. It is worth noting that the total number of students and the average number of students within each school in the 2006 data was over two times larger than those in the longitudinal dataset used in Method 3 (i.e., 17993 vs. 8785; 209 vs. 98). Because sample sizes are involved in the calculation of the reliability coefficient and residual statistics in multilevel models – larger sample sizes have higher reliability and more stable estimates on residuals – it is more reasonable to compare models with equivalent sample sizes. Therefore, a smaller sample was randomly selected from the full sample of the 2006 dataset. More specifically, a single random sample of senior students was taken, leaving about half of the students in the dataset. In this rerun of the multilevel analysis, student covariates were based on a subset of the original data. Since the institutional mean civic engagement scores among freshman students were calculated from a different group of students, those mean scores remained the same. The results from this small sample are also listed in Table 4.15 to obtain a more meaningful comparison between Methods 2 and 3.

The top panel of Table 4.15 shows the total number of students, schools, and the average number of students per school in all three samples. After random selection, Method 2 had samples sizes more similar to those of Method 3. The second panel includes the total percent of variance explained in the models and the reliability estimate on the intercept of level-1. The multilevel model used in Method 3 explained 26% of the total variance in senior students' civic engagement scores. While the model that used the

Table 4.15 Selected Results from the HLM in Method Three and Method Two

	Method 3	Method 2 (2006) Full sample	Method 2 (2006) Small sample
# of Students	8785	17993	8773
# of Schools	90	86	86
Avg. # of Students per School	98	209	102
Total Variance Explained	26%	11%	9%
Reliability Estimate on β_0	0.56	0.61	0.49
Value-added Estimates			
Min.	-4.70	-3.26	-3.74
Max.	4.29	2.20	2.77
Std. Deviation	1.33	1.10	1.12
Range of 95% CI	4.55	4.19	5.06

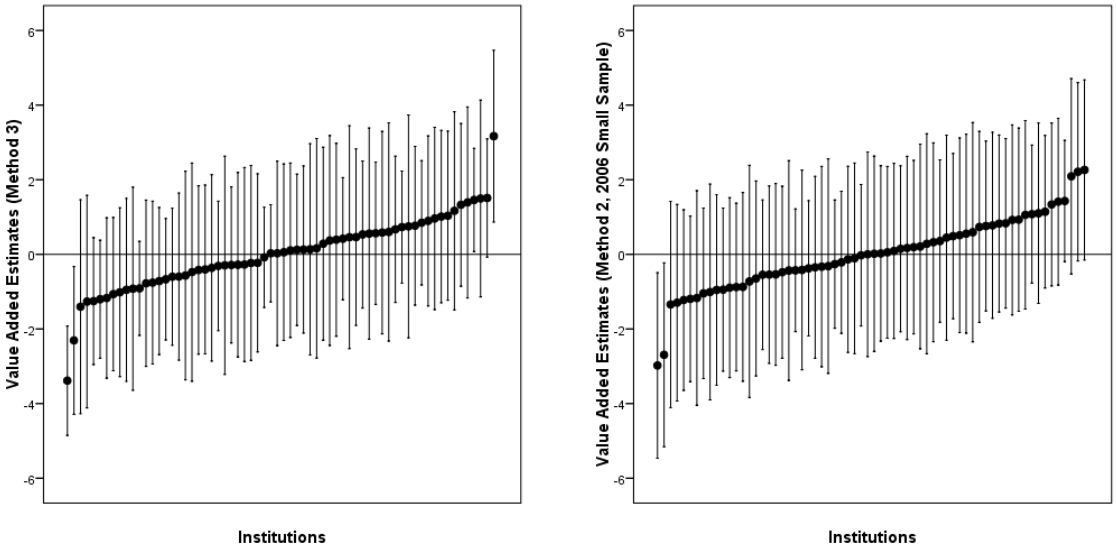
full sample in Method 2 was able to explain 11% of the total variance and the model that used the small sample was only able to explain 9% of the total variance. Value-added models aim at accounting for the differences in students' prior achievement and other background characteristics that they bring to school before receiving education there. The more variance these student covariates were able to explain, the more likely school effectiveness was to be separated out from other confounding factors. Therefore, the relatively large percent of variance accounted for by Method 3 shows a major improvement of the longitudinal method where the same student's civic engagement score in the freshman year was included in the model, rather than including aggregated civic engagement scores from a different group of students at the institutional level. For

the reliability estimate on the level-1 intercept, it was also higher in Method 3 than that in Method 2 when compared with a more equivalent sample.

The bottom panel of Table 4.15 includes selected statistics on value-added estimates (i.e., the level-2 residual u_{0j}). These residuals were standardized among all level-2 units in the sample; therefore, they have a mean of 0 which is not listed in the table. Standard errors of the value-added estimates were used to determine the width of the 95% confidence interval. The smaller the range of the 95% confidence interval is, the more accurate the estimation on the institutional value-added is. HLM has the advantage of calculating a standard error of the residual estimate (or its 95% confidence interval) for each institution. Thus, it provides more information about each institution's value-added estimate. The average range of the 95% confidence interval of the value-added estimates across all institutions is listed at the bottom of the table. Comparing Methods 3 and 2 using the small sample, Method 3 has a smaller standard error of the value-added estimates on average.

The smaller 95% confidence interval range of the value-added estimates can also be observed from Figure 4.13 which plots institutional value-added estimates among the same 66 institutions in the two samples. On average, the 95% confidence interval is narrower in Method 3 (Figure 4.13a) than the equivalent sample in Method 2 (Figure 4.13b). As shown in these figures, most 95% confidence intervals overlap with zero, indicating a majority of institutions did not differ significantly from the average institution in fostering students' development in civic engagement. Only four institutions (6%) in Method 3 were identified to be statistically different from the average. Two of

them were identified to be below the average and two were above the average. Partly due to the large standard error of the value-added estimate, Method 2 was only able to identify two institutions that were significantly below the average. Again, considering that 66 simultaneous inferences were made based on a single dataset, an average of 3.3 significant results would be found when the null hypothesis of no significant differences is true. It is hard to say whether the four institutions detected as significantly different from the average were really more (or less) effective institutions.



(a) Method Three (b) Method Two (Small Samples)

Figure 4.13 Comparison of Value-Added Estimates

Moreover, Figure 4.13a shows that the standard errors in Method 3 varied appreciably across institutions. The number of matched students within each institution was found to be associated with these standard errors. Figure 4.14 shows a clear pattern between these two factors. The standard error of the estimates quickly decreased to about

1.05 when the number of respondents within an institution increased to 100. As the number of respondents increased, the standard errors kept decreasing, only at a much slower rate.

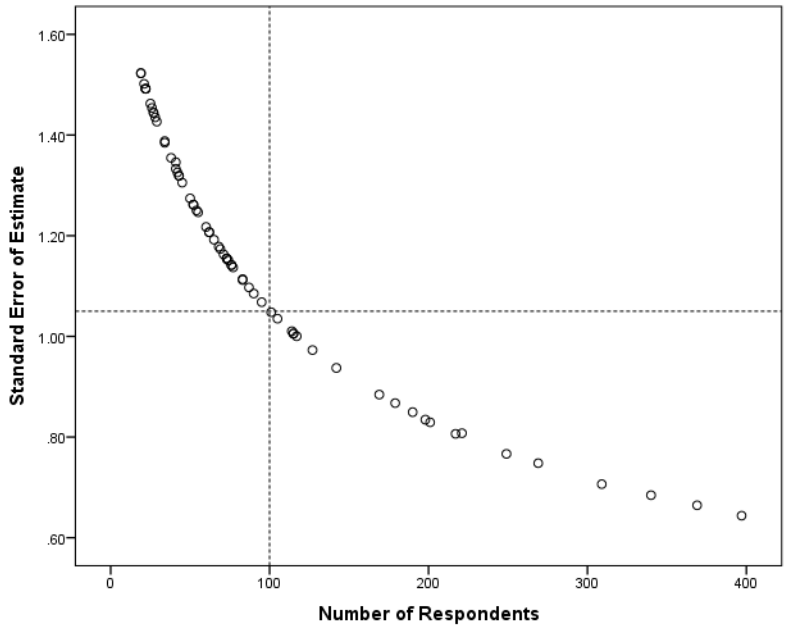


Figure 4.14 Scatter Plot for the Standard Error of Estimates and Number of Respondents in Method 3

Since it is more appropriate to compare institutions in similar peer groups, the value-added estimates for each of the nine peer groups were further examined. First, boxplots for each peer group were created based on results from Method 3. Figure 4.15 shows that although each peer group only had five to ten institutions, there was considerable variation in the value-added estimates among institutions within each peer group. Among all the peer groups, peer group 7 was the smallest peer group, consisting of five private research universities. It also had the smallest distance between the middle

two quartiles. However, one institution (#65) in peer group 7 had an extremely low value-added estimate (-4.7). One institution (#15) in peer group 8 also has a very low value-added estimate (-3.2).

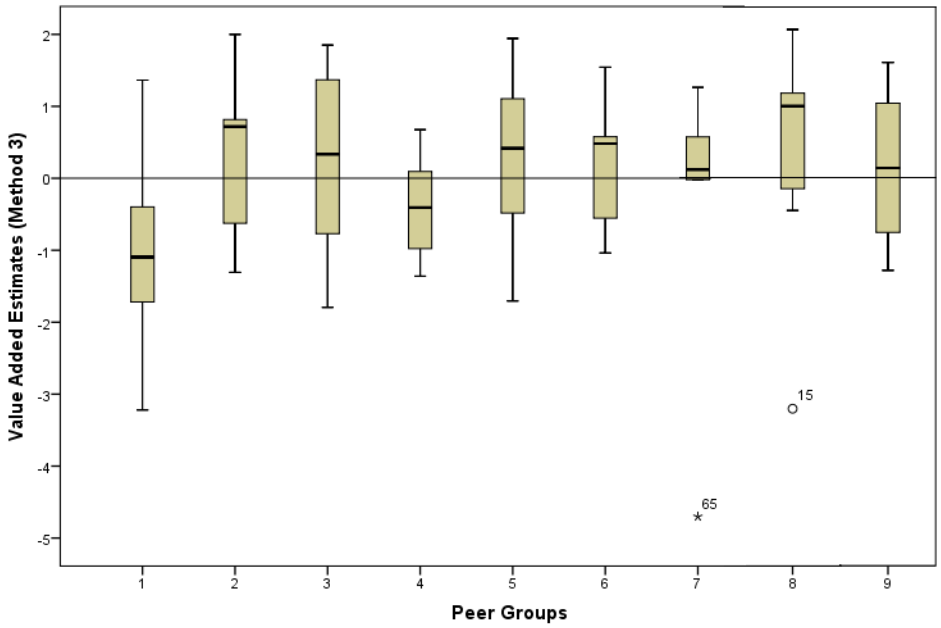


Figure 4.15 Boxplots of Value-Added Estimates by Peer Group

Next, the value-added estimates were re-standardized within each peer group with a mean of 0 and a standard deviation of 1. The 95% confidence intervals were also recalculated based on the re-standardized value-added scores. The value-added estimates for these peer groups are presented in Figure 4.16.

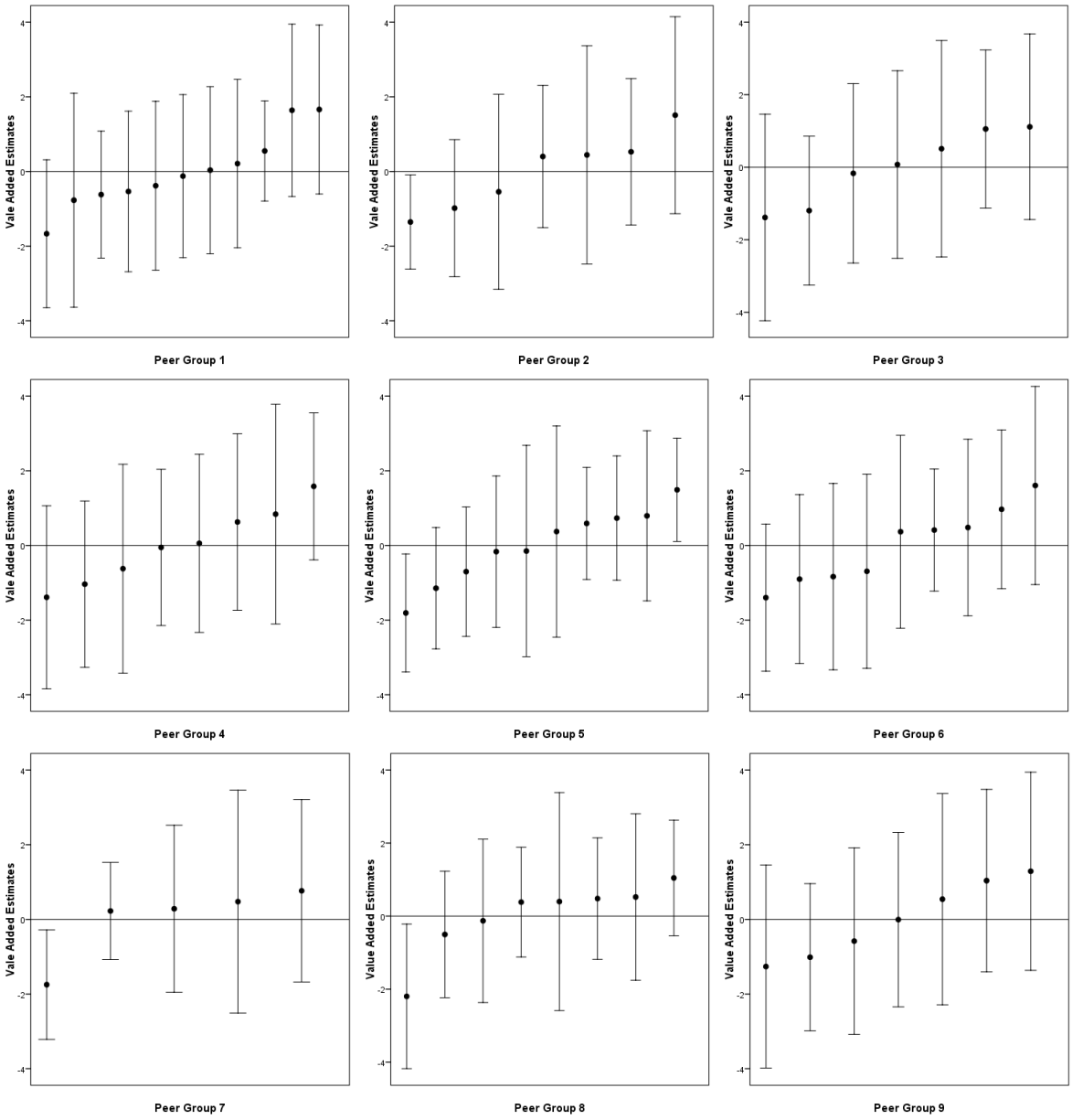


Figure 4.16 Value-Added Estimates for Nine Peer Groups

Across all nine peer groups, five institutions were found to be significantly different from the average institution in their respective peer groups. Specifically, in each of the peer groups 2, 5, 7, and 8, one institution was found to be significantly below

the average. There was also one institution found to be significantly above the average in peer group 5. Although boxplots showed nontrivial variations, most institutions were not significantly different from the average institution within each peer group with regard to fostering student development in civic engagement.

After examining each value-added method individually, the next step was to look at between-method agreement across different methods. Correlations between different models were calculated using standardized institutional value-added estimates which are listed in Table 4.16. Sixty-six institutions were included in the calculation, because some institutions in the longitudinal dataset only participated in the freshman survey in 2002 and the senior survey in 2006, but did not participate in at least one of the other two survey administrations.

Since the within-method correlations were discussed in sections 4.4.1 and 4.4.2, the focus here is on between-method correlations in Table 4.16. First, among the four correlation coefficients between analyses using Methods 1 and 2, the highest correlations occurred when the estimates were based on data from the same year. For example, when using data from 2002, the correlation between the two methods was 0.723. When using data from 2006, the correlation was 0.657. Although most correlation coefficients were statistically significant at the .05 level, when different single-year data were used, results from Methods 1 and 2 are weakly correlated. For example, the correlation between institutional value-added estimates in 2006 using Method 1 and the estimates in 2002 using Method 2 (i.e., $r = 0.235$) was not statistically significant at the .05 level ($p = .06$).

This indicates that an institution's ranking in one year using Method 1 can be completely different from its ranking in another year using Method 2.

Table 4.16 Correlations among Value-added Estimates in Three Methods

		Method 3	Method 2		Method 1	
		2006	2002	2006	2002	2006
Method 3	2006	1				
Method 2	2002	.497**	1			
	2006	.802**	.475**	1		
Method 1	2002	.362**	.723**	.246*	1	
	2006	.579**	.235	.657**	.305*	1

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

Comparing Methods 3 and 2, the between-method agreement was relatively high. The correlation was the highest when both methods were estimating institutional value added in 2006. The correlations between Methods 3 and 1 were also statistically significant, although those correlation coefficients were not as high as those between Methods 3 and 2. Again, when the value-added estimates were based on data from different years, the two methods did not agree as highly. The correlation between Method 3 in 2006 and Method 1 in 2002 was only 0.362.

In practice, when the value-added results are compared from cross-sectional and longitudinal based methods, the same year of data were used. Because Method 3 in this study was an estimation of institutional value added in 2006, the two cross-sectional methods currently used by VSA would be analogous to the models in Methods 1 and 2 based on 2006 data. The correlation coefficients between these three models are

highlighted in Table 4.16. When the same year of data was used, results from the longitudinal method (i.e., Method 3) agreed more with the HLM-based cross-sectional method (i.e., Method 2). This is represented by the highest correlation in the table (r = 0.802). The next highest correlation among the three was between the HLM-based cross-sectional method (i.e., Method 2) and the OLS-based cross-sectional method (i.e., Method 1; r = 0.657). The lowest correlation was between the longitudinal method and the OLS-based cross-sectional method (r = 0.579).

Between-method agreement was further examined in the three selected models that estimated institutional value-added in 2006. Institutional rankings from each of the three models were first categorized into four quartile groups. An index of consistency on institutional rankings was then calculated using the percent of institutions that were consistently categorized into each quartile group. These consistency indices are presented in Table 4.17.

Table 4.17 Comparisons of Quartile Group Rankings in Three Methods Estimating Institutional Value-added in 2006

	Quartile			
	1st	2nd	3rd	4th
Method 3 & Method 1	43.8%	29.4%	29.4%	50.0%
Method 3 & Method 2	50.0%	41.2%	35.3%	62.5%
Method 1 & Method 2	56.3%	35.3%	35.3%	68.8%

Similar to findings based on the correlation table, Method 3 and Method 1 had the least agreement in institutional rankings. In all four quartile groups, the index of

consistency was the lowest between Method 3 and Method 1. In the top and bottom quartile groups, Method 1 and Method 2 had the highest consistency. In the second quartile, Method 3 and Method 2 ranked schools the most consistently.

When comparing models within Method 2, increasing the number of respondents in the institution was found to be able to strengthen the associations between estimates in different years. Moreover, the number of respondents within an institution was also found to be associated with the precision of value-added estimates. For between-method comparisons, the role that the number of respondents plays was also investigated through scatterplots of institutional value-added estimates.

Figure 4.17 presents the scatterplots of institutional value-added estimates in 2006 using three different methods. Among the 66 institutions estimated in all three methods, 20 of them had at least 100 students participating in each survey administration. In Figure 4.17, these institutions with 100 or more student respondents (represented by solid black dots) were separated from those with less than 100 but more than 20 respondents (represented by circles). The dashed trend line represents the association among all institutions and the solid trend line represents the association among the institutions with at least 100 respondents. In all three graphs, the association among institutions with a large number of respondents was stronger than the association across all institutions. For example, in the scatter plot between Methods 2 and 3 (Figure 4.17c), the solid dots were more closely clustered around the trend lines than the circles.

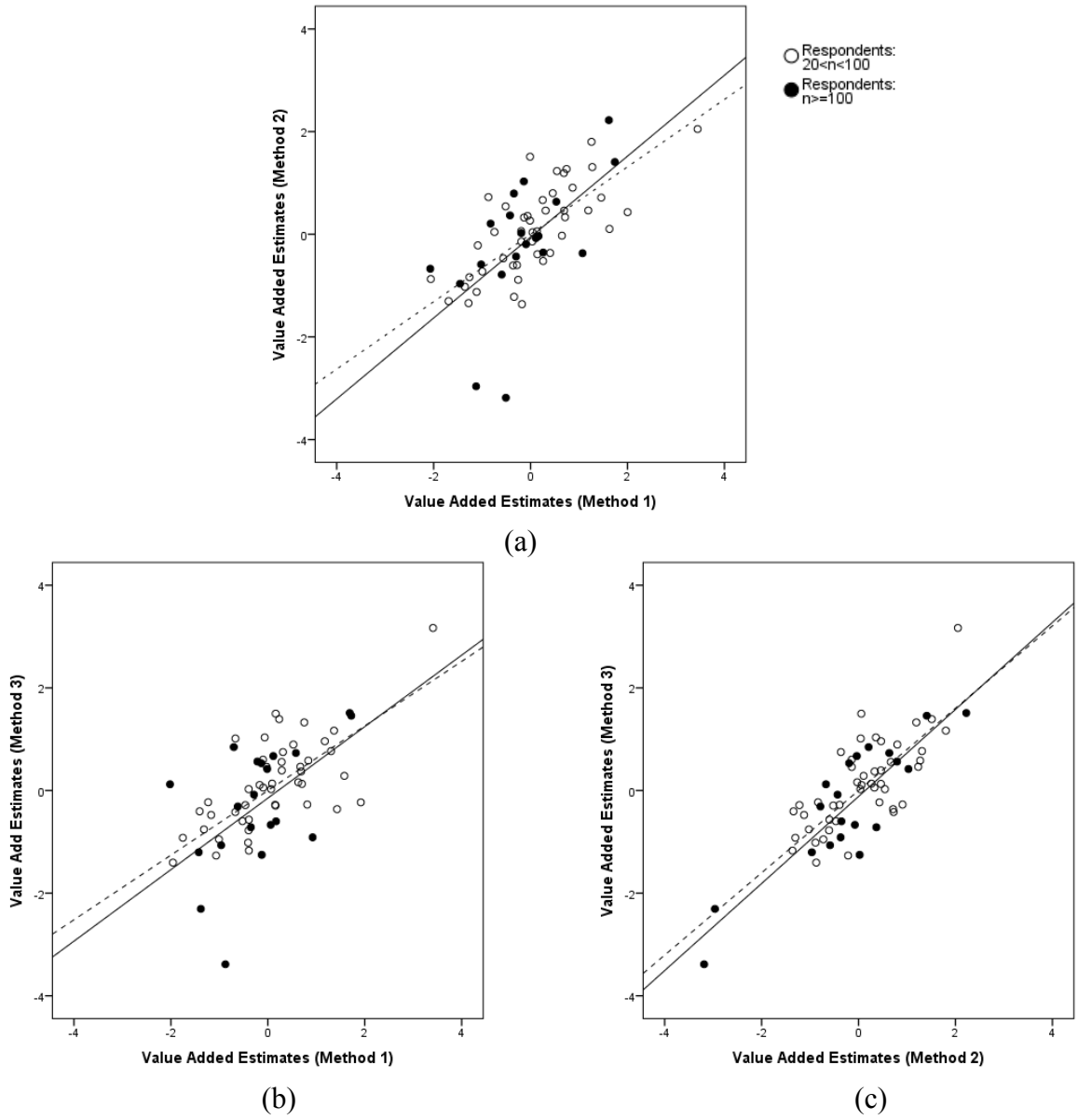


Figure 4.17 Scatterplots for Between-Method Comparisons in Three Selected Models

Finally, the three methods of estimating institutional value-added scores in 2006 were compared within peer groups. In Figure 4.16 above, institutional value-added estimates were plotted together with 95% confidence intervals. Five institutions were

found to be significantly different from the average in four of the peer group comparisons (i.e., peer groups 2, 5, 7, and 8). Figure 4.18 below presents similar plots with institutional value-added estimates in all three methods for institutions in these four peer groups. For the ease of comparison, the sequence of these institutions remained the same in each peer group. For example, the three graphs in the first row of the figure represent the value-added estimates of the same seven institutions in peer group 2 using different methods. Moreover, the institution on the far left under Method 3 is the same institution on the far left under Methods 2 and 1. Note that Methods 2 and 3 were based on HLM models, so that the 95% confidence intervals were presented as the vertical bars going through the institutional value-added estimates (i.e., the solid dots). Method 1 estimates institutional value-added scores as fixed residual values from OLS regression models. Therefore, no standard error of estimate was provided in the graphs. Two parallel dashed lines at +2 and -2 of the vertical scale shows the *Well Above Expected* and *Well Below Expected* performance levels used in VSA. In other words, when institutions had estimates outside of these reference lines, they were treated as being significantly different from the average institution in the peer group.

The first thing observed from Figure 4.18 is that some institutions estimated to be significantly below average in Method 3 cannot be identified in Method 2. The far left institutions in peer groups 2, 7, and 8 are some examples. For example, in peer group 7, the value-added estimates for the far left institution were about the same in Methods 2 and 3, but the standard error of the estimate associated with this institution was larger in Method 2. This resulted in the 95% confidence interval bar for this institution that

overlaps with zero in Method 2, so that it cannot be identified as significantly below average in peer group 7. On the other hand, some institutions were consistently identified as different from the average in both Methods 2 and 3. The first and last institutions in peer group 5 are examples in this category.

Second, using VSA's performance levels, only one institution was identified as *Well Above Expected* in Method 1. It is the institution that is the third from the left in peer group 2. The value-added estimate of this institution is +2.0. However, the value-added estimates of this institution were not significantly different from zero using Methods 2 and 3.

Third, institutional rankings within the same peer group vary across the three methods. The sequence of the institutions was sorted from low to high based on results from Method 3. This can be observed from the gradually increasing positions of solid dots in Method 3 from left to right. However, the solid dots did not follow the same monotonic increasing pattern in Methods 1 and 2. In Methods 2 and 3, the institutions ranking the highest and the lowest remained the same in three of the four peer groups, with peer group 2 being the exception. However, the sequence of institutions ranked in the middle varies considerably. In addition, the institutional ranking patterns in Methods 1 and 3 were the least consistent. In peer group 2, the two methods had different estimates for the institutions ranked the highest and the lowest. In peer group 5, the institutions ranked the lowest were also different in the two methods.

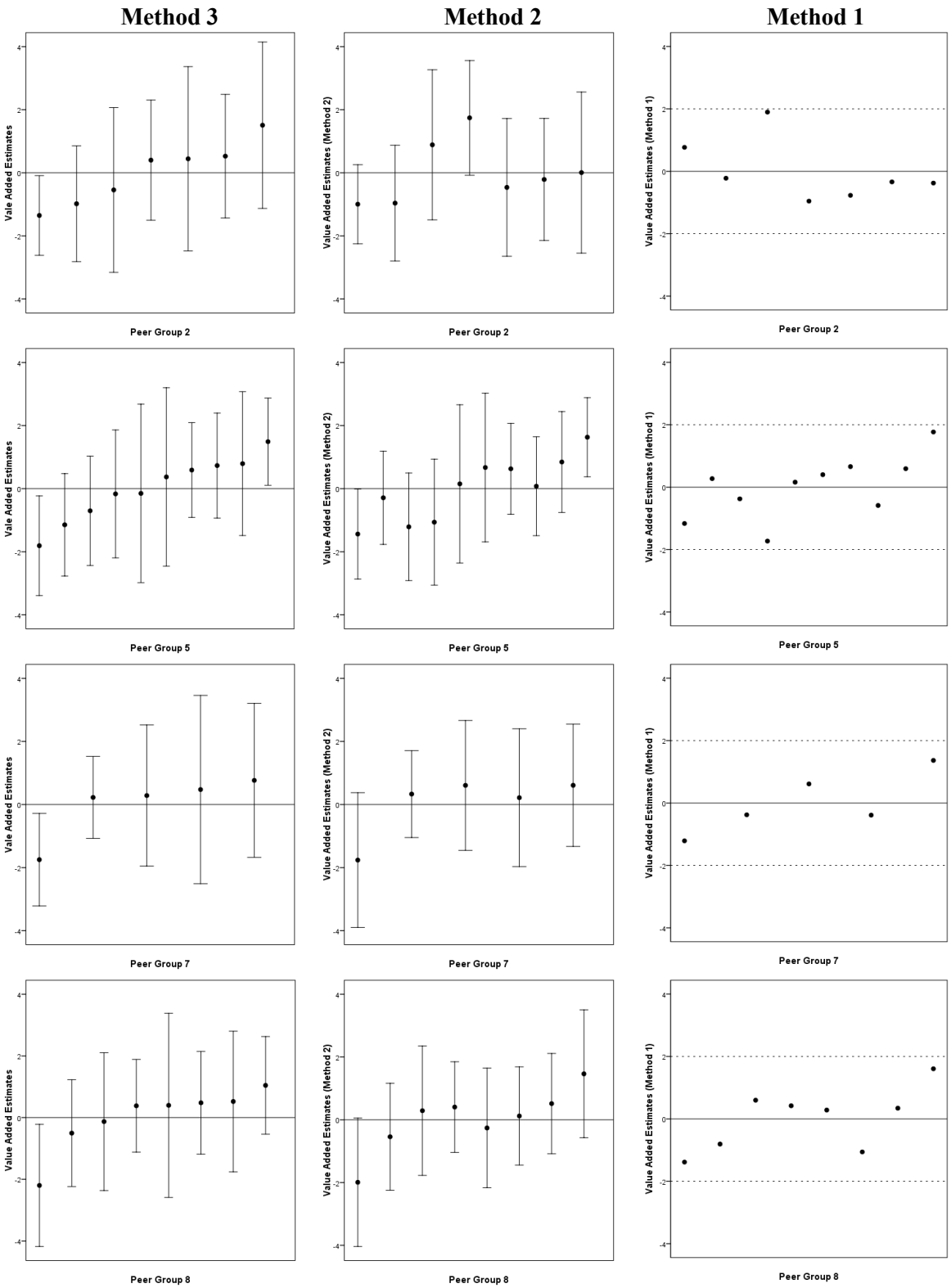


Figure 4.18 Comparison of Value-added Estimates in Three Methods for Four Peer Groups

To summarize, Method 3 analyzed longitudinal data that followed the same students at the beginning and the end of four college years. Compared with the cross-sectional model in Method 2, the model in Method 3 has shown the ability to explain a larger amount of the total variance in senior students' civic engagement scores. When compared with equivalent models with similar sample sizes, the model in Method 3 also showed higher reliability and smaller standard error estimates, which is critical in deciding whether an institution's value-added estimate is significantly different from the average. Moreover, correlations among all methods were summarized. The correlations between the two cross-sectional methods VSA currently adopts (i.e., Method 1 using 2006 data and Method 2 using 2006 data) and the longitudinal method (i.e., Method 3) were 0.58 and 0.80, respectively. The inconsistent value-added estimates were also shown in inconsistent institutional quartile groups using different methods.

Moreover, institutional value-added estimates in different methods were compared within nine peer groups. Few institutions were identified as significantly different from the average in helping students develop in civic engagement. Some inconsistencies exist between Methods 2 and 3. Furthermore, using Method 1 and the *Well Above Expected* and *Well Below Expected* performance levels set by VSA, the method consistency was even less.

Even when comparing value-added estimates within small peer groups in different methods, institutions' ranking patterns were far from identical, especially for institutions that had mid-range value-added estimations. Moreover, using Method 1, value-added estimates can be standardized and categorized into below and above expected groups.

But the OLS regression does not provide an indicator of precision of these value-added estimates. Because hierarchical linear models have the advantage of providing standard errors of the value-added estimates, institutions can be identified as significantly below or above average for a certain confidence level. This is one important improvement in using the HLM analysis in value-added estimations.

CHAPTER 5. CONCLUSIONS

5.1 Summary of Findings

To better understand whether and how undergraduate students learn and progress, multiple efforts have been made both to measure higher education outcomes and to estimate value-added in higher education institutions. As the last stage of formal education for most Americans, four year colleges educate future citizens. They play a critical role in preparing young adults to become responsible citizens in a democratic society and to become knowledgeable and skillful individuals in the workplace. Although there is agreement that higher education outcomes are multidimensional, current efforts mainly focus on measuring traditional cognitive educational outcomes such as writing, reasoning, and critical thinking. Moreover, despite the call to treat the longitudinal pretest-posttest design as the "gold standard" in estimating college students' learning in the absence of randomized experiments (Seifert, Pascarella, et al., 2010), the most influential value-added study in higher education, VSA, still adopts a cross-sectional design because of practical considerations. In this context, the present study extends current efforts in measuring higher education outcomes and exploring the differences among three value-added methodologies.

This investigation focused on a particular noncognitive outcome, civic engagement. Based on undergraduate survey data, a three-stage study was first carried out to develop a quantitative scale representing students' attitudes towards civic engagement. Then, different hierarchical linear models were adopted to estimate associations between undergraduate students' civic engagement scale scores in the senior

year and student and institutional characteristics, after adjusting for students' civic engagement scores in the freshman year. Similarities and differences among cross-sectional and longitudinal value-added models in estimating institutional value-added were examined. The following section summarizes findings from the development of the civic engagement scale, as well as three guiding research questions of the study.

5.1.1 Associations between Civic Engagement and Student-level Variables

High quality measurements are critical for subsequent statistical analysis. Before carrying out the HLM analyses, an 8-item civic engagement scale was developed through a three-stage study which has demonstrated good psychometric properties for the civic engagement scale. First, consistent with previous studies, the civic engagement construct measured in the CIRP surveys was found to be multidimensional. A two-factor solution emerged, in which all items had relatively high loadings (0.5 to 0.8) on the first derived factor. The 8-item factor is related to the attitudes and beliefs students have towards civic engagement. Further exploratory factor analysis and reliability analysis show that the factor is unidimensional and has high reliability coefficients (around 0.84). Second, CFA was used to validate the factor structure of this 8-item scale. Multiple goodness-of-fit indices provided support that the hypothesized 8-item scale for civic engagement had a moderate to good fit with the data. Next, multisample CFA was used to assess the measurement invariance of the 8-item scale. Results showed that both factor patterns and factor loadings were consistent across different CIRP surveys and administrations. This finding supports the hypothesis that the civic engagement scale measures the same

construct across the four CIRP surveys, which were administered to students at different postsecondary levels and in different years. Finally, the Rasch partial credit model was used to derive the scale scores. Results showed appropriate distributions on item difficulties with regard to students' attitudes towards civic engagement. A good model-data fit was also found across all eight items.

In the HLM analysis, the 2002-2006 longitudinal dataset in which institutions had at least 20 students was first analyzed. Several student-level variables were identified as being associated with students' civic engagement scores in the senior year, even after adjusting for their freshman civic engagement scores. Among student demographics and family background variables, students' race/ethnicity was the only one found to be associated with civic engagement scores in the senior year. In comparison to students with similar background characteristics, Hispanic students scored significantly higher on civic engagement in the senior year than their white peers.

Moreover, all other student-level variables examined in the study were significantly associated with senior students' civic engagement scores. They are: students' educational aspirations, SAT/ACT scores, high school GPA, college GPA, and civic related activities during college. In general, when a student aspired to earn a graduate degree (e.g., master's degree, law degree, divinity degree, Ph.D./Ed.D. degree) rather than stopping with a baccalaureate degree, the student was estimated to score higher in senior year civic engagement after accounting for their freshman civic engagement scores and other covariates in the model. In addition, participation in civic related activities, such as performing volunteer work, voting in student elections, and

participating in student clubs/groups, were positively associated with students' civic engagement scores. An interesting finding is that although college GPA was positively associated with students' civic engagement scores after controlling for all other covariates in the model, high school GPA and SAT/ACT scores were both negatively associated with these scores. However, the statistical significance does not necessarily mean the results are important or meaningful. Actually, the changes in senior students' civic engagement scores are so small that they do not indicate a practical significance.

Using latent class analysis, a student profile variable was constructed which categorized undergraduate students into five groups – religious oriented, status strivers, artistes, hedonists, and disengaged. These student profile groups generally correspond with college student profile classifications in previous studies (Astin, 1993b; Clark & Trow, 1966; Kuh et al., 2000). In the HLM analysis, disengaged students were found to score significantly lower in civic engagement at the senior year than their peers with similar freshman civic engagement scores and similar backgrounds. On average, disengaged students scored the lowest in civic engagement both in the freshman year and in the senior year. This is also consistent with previous research findings (e.g., Kuh et al., 2000) that disengaged students had below average self-reported gains in intellectual development, general education, personal development, life-long learning, etc.

These results were further confirmed in a replication in which only institutions with larger student samples of at least 100 student respondents were included in the analyses. Most variables were statistically significant at the .01 level in both models.

5.1.2 Associations between Civic Engagement and School-level Variables

Despite the fact that the between school variance component was only three percent of the total variance in the civic engagement measure, several institutional characteristics were associated with civic engagement. Using the basic Carnegie classification (see footnote 19 for detailed definitions of basic Carnegie classifications), senior students from master's colleges and universities demonstrated a significantly higher value on civic engagement than those from baccalaureate colleges. Moreover, senior students from undergraduate instructional programs categorized as Arts & Sciences plus professions scored higher on civic engagement than those from balanced Arts & Sciences and professions programs (see footnote 20 for detailed definitions of Carnegie classifications in undergraduate instructional programs). Institutions with professions plus Arts & Sciences instructional programs had the lowest civic engagement scores both at the freshman year and at the senior year among all institutions. These results agree with findings at the student-level that students who plan to pursue advanced degrees such as a Ph.D./Ed.D. or a law degree are predicted to score higher on civic engagement.

Regional location was also associated with students' development in civic engagement. Students from institutions located in the Far West scored significantly higher than those from institutions located in the Mid-Atlantic area in senior civic engagement scores, after accounting for freshman civic engagement scores and other background characteristics. On average, students from institutions located in the Mid-

Atlantic area also had the lowest civic engagement scores both at the freshman year and at the senior year.

5.1.3 Comparison of Three Value-added Methods

A major research question raised in this study is whether methodology matters in calculating institutional value-added. Using data from CIRP surveys, three value-added methods were compared. Namely, an OLS-based cross-sectional method (Method 1), an HLM-based cross-sectional method (Method 2), and an HLM-based longitudinal method (Method 3) were used to estimate relative institutional effectiveness with regard to civic engagement. Although HERI administers a series of three student surveys at the freshman year, after the first college year, and at the senior year, only 2-wave data from freshman and senior students are available in this study given the limited number of respondents in the Your First College Year Survey.

In Method 1, the institution is the unit of analysis. Institutional value-added is calculated as a difference in differences score. It is the difference between two residual scores in separate OLS models predicting institutional mean civic engagement scores among freshman and senior students. In Methods 2 and 3, HLM-based models are used and the student is the unit of analysis. The institutional value-added estimates are differences between the observed and expected senior students' civic engagement scores after adjusting for students' civic engagement in the freshman year and other covariates. Therefore, strictly speaking, the resulting value-added scores in this study are adjusted

changes in status scores with regard to civic engagement, rather than differences in the institutional growth rates on civic engagement.

After fitting different institutional value-added models, their similarities and differences were first demonstrated through within-method stability and between-method agreement. Since Methods 1 and 2 were used to estimate institutional value-added in both 2002 and 2006, the within-method stability was measured by the Pearson correlation coefficient between the two estimates. A higher year-to-year consistency was found in Method 2 than that in Method 1 (i.e., 0.50 vs. 0.30). However, both within-method stabilities are relatively weak and institutional rankings in different years fluctuate dramatically. This could be partly due to the true changes over four years in institutional effectiveness. However, the volatility may also result from other sources. For example, students self-selected to participate in the CIRP surveys. Most institutions recruited a small group of students whose civic engagement levels were not representative of the entire institution. In other words, the student sample obtained was an under-representative sample of the student cohort in most institutions. Moreover, under the cross-sectional study design, information from two different student cohorts – the freshman and the senior classes – were collected and the civic engagement scores of the freshmen were used as a proxy of those who matriculated to the institution four years earlier. Therefore, not only was each of the freshman and senior student samples under-representative of the entire student cohort in that institution, these two student samples were also different in nature. They could be under-representative and incomparable in different ways, so that the freshman student sample should not be used in place of the

actual freshmen students four years earlier. All these factors can introduce selection bias and sampling errors to the value-added estimates. Because many factors, including instrument, measurement, and analytical issues, could confound the true value-added estimate (Braun et al., 2010), multiple years of data should be used before making high-stakes decisions based on these institutional value-added scores.

The number of respondents within institutions was found to play a role in model estimation in Methods 2 and 3, but not in Method 1. In Method 2, when the minimum number of students within each institution increased from 20 to 100, the within-method stability increased. Moreover, increasing the number of minimum respondents resulted in a decrease in the standard errors of the value-added estimates provided in both HLM-based methods. Although a large sample size does not guarantee representativeness, this finding should motivate institutions to recruit as many respondents as possible.

It is important to bear in mind that the comparisons in institutional rankings and correlations among different value-added estimates cannot alone determine the superiority of a particular method. The primary goal of this phase of the study was to document differences among these different methods. In the examination of between-method agreement, results from the HLM-based longitudinal method (Method 3) are more similar to those from the HLM-based cross-sectional method (Method 2) than those from the OLS-based cross-sectional method (Method 1). The two HLM-based methods estimating institutional effectiveness in 2006 had the highest correlation (r = .80) among all pairs of comparisons. However, the volatility among institutional rankings and institutions' quartile groups when using different HLM-based methods is still a problem.

Results further show that Method 3 was able to explain almost three times of the total variance in senior students' civic engagement scores over an equivalent model in Method 2 (26% vs. 9%). On average, the standard error (or the 95% confidence interval) of the value-added estimates in the VAM using longitudinal data was also smaller than that when cross-sectional data was used. Narrower 95% confidence intervals demonstrate an advantage of the longitudinal method over the cross-sectional method, which models a proxy of students' prior achievement at the institutional level. A special note was further made on the problem of multiplicity or simultaneous inferences. Caution is needed before drawing conclusions on institutions which emerged as significantly below or above the average because of the multiple inferences made on the single dataset.

In the end, the three sets of institutional value-added estimates were compared within each of the nine peer groups. The sequence of institution rankings within these small peer groups still fluctuate considerably across three different methods. Across all peer groups, five institutions were identified by Method 3 as significantly different from average institutions. Method 2 only identified two of them. Method 1 failed to identify any of those institutions.

5.2 Policy Implications

Most institutions aim to cultivate well-rounded citizens who are intellectually curious and socially responsible. Consequently, when evaluating the quality of higher education institutions, both cognitive and noncognitive aspects of higher education outcomes should be considered, and the role of higher education in developing well-

rounded citizens needs to be acknowledged. VSA has put great effort into measuring cognitive outcomes, such as critical thinking, analytic reasoning, and written communication, and estimating institutional value-added with regard to these measures. However, noncognitive outcomes are missing from their accountability system. Although challenges exist in identifying critical noncognitive outcomes and developing credible measures of these outcomes, noncognitive outcomes are indispensable for obtaining a complete picture of the quality of higher education institutions.

Civic engagement is a critical constituent of the set of noncognitive outcomes in higher education. Recognizing the importance of having engaged citizens in a democratic society, American colleges and universities, regardless of their diverse education missions, all have a common goal that speaks to the cultivation of active and civically engaged citizens (Ehrlich, 2000; Pew Partnership for Civic Change, 2004). In AAC&U's (2012) recently released report, *A Crucible Moment: College Learning & Democracy's Future*, "anemic US civic health" was documented, and a national call to action was made on civic learning and democratic engagement. It is a critical time to embrace measures to improve civic engagement in all levels of education, especially higher education.

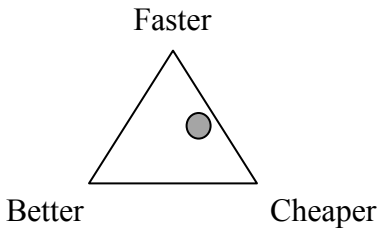
With regard to value-added methodologies, it is important to take into consideration that different value-added methods, different models with variations using the same method, and the same model using data from different years can produce substantially different results. This study used the same covariates in all three models in order to make different value-added methods comparable. Different institutional value-

added estimates were produced using data from the same year. Institutional rankings fluctuate across the entire range of institutions, including those ranked at the top and at the bottom. When comparing institutions within small and homogenous peer groups, institutional rankings still varied considerably. Because institutional rankings changed so frequently, this can create great uncertainty, and potential injustice, if these value-added results are used in making high-stakes decision on school rewards or sanctions.

True value-added models hold the promise of isolating the relative contributions of individual institutions, or teachers and programs. They require longitudinal data obtained from the same students over time, providing better statistical control over students' pre-measures on the outcome. As a result, these longitudinal models account for more variance in the outcome measure and typically yield value-added estimates that are both more accurate and more precise. Conversely, cross-sectional designs have clear advantages of reducing costs in data collection and providing results more quickly. However, with the use of a proxy for students' prior achievement, cross-sectional value-added models are only pseudo-VAMs. Their fragility lies in the use of imprecise proxies which in turn add confounding errors to the value-added estimates.

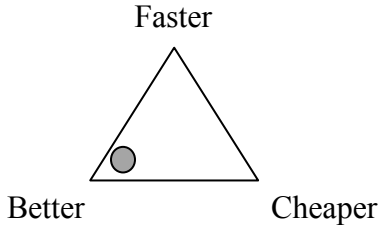
As discussed in chapter two, the debate between longitudinal and cross-sectional methodologies seems to have already been settled in favor of longitudinal methods. If the differences between the two methods were found to be trivial, concerns of cost and ease of administration might trump the desire to use true VAMs. However, the differences are not trivial, as this study demonstrated. Therefore, the approach of avoiding longitudinal research designs because they are more difficult to administer

needs to be reconsidered. Giving up accurate estimates in a longitudinal design for a cross-sectional design because of considerations of cost and effort is not a good trade-off, especially when high-stakes decisions are made on the basis of these value-added estimates. As Chun (2006) demonstrated in the "iron triangle," with various practical limitations, striking a balance between *faster*, *cheaper*, and *better* is crucial. When *faster* and *cheaper* is chosen over *better*, "one must be willing to accept that the process might not necessarily be *better*" (p.1):



However,

> Quality should never be negotiable. If one completes analyses cheaper and faster, but there are no meaningful conclusions that can be drawn, the entire exercise is rather worthless. Having better data, conducting better analyses, and coming to better conclusions should be the key… Until the day comes when we can choose three, perhaps we should instead start by choosing one: (p.6)



There is no shortcut to collecting high quality data in a manner that minimizes confounding factors in estimating institutional value-added. Continuing to use methods to cut corners will result in more biased results, which could lead to inaccurate and distorted estimates of institutional effectiveness and other unintended consequences.

Results from the first two research questions of the study also have policy implications. In the first two research questions, several student- and institution-level covariates were found to be associated with senior students' civic engagement level after adjusting for their civic engagement in the freshman year. Disengaged students, those with low college GPAs, and those who are rarely involved in student clubs, volunteer work, or voting in student elections should be the target of school professionals to help them develop positive attitudes towards civic engagement. Especially since students' attitudes towards civic engagement vary more greatly within institutions than between institutions, the focus should be put on helping these high-risk students in every single institution. Moreover, degree granting characteristics and the location of an institution are part of the institutional context that is exogenous to instructional practices. Nonetheless, attention should be given to these institutions and specific types of instructional programs to promote their students' and future adult citizens' civic engagement level.

5.3 Limitations

There are a number of limitations to this study. First, a noncognitive scale was derived and used in subsequent value-added analyses. During the construction of the civic engagement scale and model building process, multiple attempts were made to refine the analysis – including the use of exploratory and confirmatory factor analysis on item selections, the use of Rasch analysis to create the scale scores, and the adjustment on measurement errors. However, the reliability of a noncognitive instrument is normally lower than that of a cognitive assessment due to the number and type of items included. Thus, limitations in the precision and sensitivity of the scale in detecting differences in students' attitudes towards civic engagement are critical factors to consider before interpreting the value-added results.

In addition, the CIRP surveys are not designed specifically to measure the level of undergraduate students' civic engagement. The derived scale was only able to reliably capture the inward aspect of the multidimensional civic engagement construct. The outward aspect of students' participation in civic related activities was not included in the civic engagement scale. Therefore, the civic engagement scale developed in this study was only a subscale of the multidimensional civic engagement construct. The value-added estimates were the relative institutional effectiveness with regard to developing students' positive attitudes and beliefs towards civic engagement. Construct underrepresentation may become an issue had the findings been interpreted as a comprehensive estimate of institutional effectiveness on the multidimensional civic engagement construct.

Thirdly, because undergraduate students participated in the CIRP surveys on a voluntary basis, non-random samples are used in estimating institutional value-added scores. For example, some campuses use various incentives to recruit students in the survey administration, while others do not. Students with different characteristics may be attracted to complete the survey on different campuses. These non-random samples together with the relatively small samples available in some institutions may contribute bias to the study results. Although post-stratification weights were used and models were replicated within institutions with larger sample sizes, these statistical adjustments cannot completely compensate for the fact that these students are likely not representative of the entire populations of the participating institutions.

In general, these limitations are related to the nature of the data analyzed in this study. Student data were collected from non-experimental survey studies in which students were not randomly assigned to postsecondary institutions. Without randomization, institutional effectiveness could be confounded with selection bias; the contributions of institutions to students' learning cannot be completely isolated. Therefore, causal inferences cannot be made to link student outcomes with institutional effectiveness with regard to civic engagement, and, consequently, the resulting institutional value-added estimates must be interpreted with caution.

5.4 Modeling Student Outcomes in Higher Education – Looking Forward

There has been considerable demand from various stakeholders to undertake the measurement of undergraduate students' learning outcomes and employ these outcomes in the evaluation of higher education institutions. However, much more research is needed before higher education outcomes can be properly modeled and used for accountability purposes. The first area of research needed is in obtaining greater consensus on the critical student outcomes in higher education. This study has shed some light on the value and utility of measuring noncognitive outcomes such as civic engagement. Other important higher education outcomes, both cognitive and noncognitive, need to be identified to fully capture institutional effectiveness. Given the diversity of higher education institutions, different key outcomes may need to be developed based on the unique mission and values of each individual institution. This will be a challenging but rewarding task.

Second, measuring noncognitive student outcomes for use in estimating value-added of higher education institutions is a relatively new research area. Most studies in higher education rely on survey instruments such as CIRP or NSSE, two comprehensive surveys that measure undergraduate students' college experiences and student engagement. However, these surveys are not developed as all-purpose measures. In future research, specific noncognitive measures need to be developed with a focus on measuring certain higher education outcomes.

Third, longitudinal study designs that follow students more than two times and continue into their adult lives are worth exploring. As Bennett (2001) argues, alumni,

rather than graduating seniors, need to be assessed to properly measure quality in higher education, since it can take years to reveal some of the effects of a college education. Most longitudinal studies in higher education currently adopt a typical two-wave pre-post design, collecting information from freshmen and seniors. Students' long term development will be better revealed when information is collected at three or more time points.

Fourth, value-added models vary with respect to controlling for different covariates at different levels of the model and using multiple years of data. Using the longitudinal data available, this study only examined one value-added model and estimated institutional effectiveness using a single student cohort. Year-to-year consistency was not investigated. Future studies may examine the year-to-year consistency of this value-added model or explore other forms of value-added models with multiple student cohorts. Depending on the nature of the student outcome measured, some value-added models may be found to perform better than the others. Moreover, given the available information, students who transferred during the four college years were not included in the longitudinal study. Future studies may consider models that are specifically developed to account for cross classification of students and schools to further explore this practical issue.

Fifth, beyond receiving a quantitative measure of institutional value-added, educators and administrators are looking for operational suggestions regarding instructional practices and institutional management approaches to assist students to better learn and develop. Therefore, institutional value-added estimates only act as a

driving force for institutional improvement. To inform teaching and learning, further qualitative studies are needed to investigate the complexity of undergraduate students' experiences and higher education outcomes.

Finally, the entire idea of accountability in higher education needs careful consideration. Current initiatives from collecting comparable consumer information across different institutions, to measuring cognitive higher education outcomes, to estimating institutional value-added are important steps to advance higher education. However, much larger steps are needed before linking estimated institutional effectiveness with institutional accountability. The accountability system itself needs future validation, which involves addressing various challenges with instruments, measurements, and analytic issues and collecting evidence on the interpretations and consequences of the accountability system. As Shavelson (2010) argues, accountability of individuals and institutions is an essential component of democratic society. But in the higher education context, it has the potential to be both a powerful tool for institutional improvement and an instrument which brings unintended negative consequences. At this stage, the focus on accountability should shift to providing legitimate consumer information and using institutional value-added estimates as a driving force for institutional improvement. In this way, the small step achieved so far will continue to move forward in a promising direction.

REFERENCES

Abe, J. A. (2005). The predictive value of the Five-Factor Model of personality with preschool age children: A nine year follow-up study. *Journal of Research in Personality, 39*, *423*-442.

ACT. (2009). *ACT report: Voluntary System of Accountability learning gains methodology*. Retrieved May 10, 2010 from http://www.voluntarysystem.org/docs/cp/ACTReport_LearningGainsMethodology.pdf

Adelman, C., Ewell, P., Gaston, P., & Schneider, C. G. (2011). *The degree qualifications profile*. Indianapolis, IN: Lumina Foundation for Education.

American College Personnel Association. (1994). *The student learning imperative: Implications for student affairs*. Washington, DC: Author.

American Council of Trustees and Alumni. (2007). *Why accreditation doesn't work and what policymakers can do about it: A policy paper from the American Council of Trustees and Alumni*. Washington, DC: Author. Retrieved Dec. 25th, 2010 from goacta.org/publications/downloads/Accreditation2007Final.pdf

Anderson, J. M. (2011, July 5). *Beware the "outcomes and assessment" movement*. Center for the American University at the Manhattan Institute. Retrieved September 9, 2011 from http://www.mindingthecampus.com/originals/2011/07/_for_yet_another_glimpse.html

Artelt, C., Baumert, J., Julius-McElvany, N., & Peschar, J. (2003). *Learners for life: Student approaches to learning. Results from PISA 2000.* Paris: Organisation for Economic Co-operation and Development.

Arum, R., & Roksa, J. (2011). *Academically adrift: Limited learning on college campuses*. Chicago and London: University of Chicago Press.

Association of American Colleges and Universities. (2002). *Greater expectations: A new vision for learning as a nation goes to college*. Washington, DC: Author.

Association of American Colleges and Universities. (2004). *Taking responsibility for the quality of U.S. baccalaureate degree: A report from the Greater Expectations Project on Accreditation and Assessment.* Washington, DC: Author.

Association of American Colleges and Universities. (2005). *Liberal education outcomes: A preliminary report on student achievement in college*. Washington, DC: Author.

Association of American Colleges and Universities. (2007). *College learning for the new global century*. Washington, DC: Author.

Association of American Colleges and Universities. (2010). *Civic Engagement VALUE Rubric*. Washington, DC: Author.

Association of American Colleges and Universities. (2011). *VALUE: Valid Assessment of Learning in Undergraduate Education – VALUE rubrics*. Washington, DC: Author.

Association of American Colleges and Universities. (2012). *A crucible moment: College learning & democracy's future*. Washington, DC: Author.

Astin, A. W. (1970). The methodology of research on college impact, Part I and Part II. *Sociology of Education, 43*(4), 223-254, 437-450.

Astin, A. W. (1973). Measurement and determinants of the outputs of higher education. In L. Solmon & P. Taubman (Eds.), *Does college matters? Some evidence on the impacts of higher education* (pp. 107-127). New York: Academic Press.

Astin, A. W. (1993a). *What matters in college? Four critical years revisited*. San Francisco: Jossey-Bass.

Astin, A. W. (1993b). An empirical typology of college students. *Journal of College Student Development, 34*(1), 36-46.

Astin, A. W. (1996). Involvement in learning revisited: Lessons we have learned. *Journal of College Student Development, 37*, 123-134.

Astin, A. W. (2003). Studying how college affects student: A personal history of the CIRP. *About Campus, 8*(3), 21-28.

Astin, A. W., & Lee, J. J. (2003). How risky are one-shot cross-sectional assessments of undergraduate students? *Research in Higher Education*, *44*(6), 657-672.

Astin, A. W., & Sax, L. J. (1998). How undergraduates are affected by service participation. *The Journal of College Student Development, 39*(3), 251-263.

Astin, A. W., Vogelgesang, L. J., Ikeda, E. K., & Yee, J. A. (2000). *How service learning affects students*. Los Angeles: University of California, Higher Education Research Institute.

Astin, H. S., & Antonio, A. L. (2004). The impact of college on character development. *New Directions for Institutional Research, 122,* 55-64.

Banta, T. W. (2006). Reliving the history of large-scale assessment in higher education. *Assessment Update , 18*(4), 3-4, 15.

Banta, T. W. (2007, January 26). A warning on measuring learning outcomes. *Inside Higher Education*. Retrieved April 22, 2011 from http://www.insidehighered.com/views/2007/01/26/banta

Banta, T. W. (2008). Editor's notes: Trying to clothe the emperor. *Assessment Update, 20*(2), 3-4, 15-16.

Banta, T. W., & Pike, G. R. (2007). Revisiting the blind alley of value-added. *Assessment Update , 19*(1), 1-2, 14-15.

Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44*, 1-26.

Bauer, K. W. (2004). Conducting longitudinal studies. *New Directions for Institutional Research, 121,* 75-90.

Baxter Magolda, M. B. (1999). Defining and redefining student learning. In E. Whitt (Ed.), *Student learning as student affairs work* (pp. 35-49). Washington, DC: National Association of Student Personnel Administrators.

Baxter Magolda, M. B. (2001). *Making their own way: Narratives for transforming higher education to promote self-development*. Sterling, VA: Stylus Publishing.

Benjamini, Y. & Braun, H. I. (2002). John W. Tukey's contributions to multiple comparisons. *The Annals of Statistics, 30*(6), 1576-1594.

Bennett, D. C. (2001). Assessing quality in higher education. *Liberal Education, 87*(2), 40-45.

Berkner, L., He, S., & Cataldi, E. F. (2002). *Descriptive summary of 1995–96 beginning postsecondary students: Six years later* (NCES 2003–151). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Beyer, C. H., Gillmore, G., & Fisher, A. T. (2007). *Inside the undergraduate experience: The University of Washington's study of undergraduate learning*. Bolton, MA: Anker Publishing Company, Inc.

Bok, D. (2006). *Our underachieving colleges: A candid look at how much students learn and why they should be learning more*. Princeton, NJ: Princeton University Press.

Boston College. (2007). *The journey into adulthood: Understanding student formation*. Chestnut Hill: Boston College. The office of marketing communications. OMC 2829.

Bowen, H. R. (1977). *Investment in learning: The individual and social value of american higher education*. San Francisco: Jossey-Bass.

Bowles, S., Gintis, H., & Osborne, M. (2001). The determinants of earnings: A behavioral approach. *Journal of Economic Literature, 39*(4), 1137-1176.

Bowman, N. A. (2011). Promoting participation in a diverse democracy: A meta-analysis of college diversity experiences and civic engagement. *Review of Educational Research, 81*(1), 29-68.

Boyer, S. P., & Sedlacek, W. E. (1988). Noncognitive predictors of academic success for international students: A longitudinal study. *Journal of College Student Development, 29*, 218-222.

Braun, H. I. (2005). *Using student progress to evaluate teachers: A primer on value-added models*. Princeton, NJ: Educational Testing Services.

Braun, H. I., Chudowsky, N., & Koenig, J. (Eds). (2010). *Getting value out of value-added: Report of a workshop.* Washington, DC: The National Academies Press.

Braun, H. I., Kirsch, I., & Yamamoto, K. (2011). An experimental study of the effects of monetary incentives on performance on the 12th-grade NAEP reading assessment. *Teachers College Record, 113*(11), 2309-2344.

Briggs, D. C., & Weeks, J. P. (2009). The impact of vertical scaling decisions on growth interpretations. *Educational Measurement: Issues and Practice, 28*(4), 3-14.

Business-Higher Education Forum. (2003). *Building a nation of learners: The need for changes in teaching and learning to meet global challenges*. Washington, DC: Author.

Callan, P. M., & Finney, J. E. (2002). Assessing educational capital: an imperative for policy. *Change, 34*, 25-31.

Campbell, D. T. (1975). Assessing the impact of planned social change. In G. M. Lyons (Ed.), *Social research and public policies: The Dartmouth/OECD Conference* (pp.35). Hanover, NH: Public Affairs Center, Dartmouth College.

Caprara, G. V., Barbanelli, C., Pastorelli, C., Bandura, A., & Zimbardo, P. G. (2000). Prosocial foundations of children's academic achievement. *Psychological Science, 11,* 302-306.

Carnegie Foundation for the Advancement of Teaching. (n.d.a). *Classification Description*. Retrieved December 4th, 2011 from http://classifications. carnegiefoundation org/descriptions/basic.php

Carnegie Foundation for the Advancement of Teaching. (n.d.b). *Classification Description: Undergraduate Instructional Program Classification*. Retrieved December 4th, 2011 from http://classifications.carnegiefoundation.org/ descriptions/ugrad_program.php

Case, C. E., & Greeley, A. M. (1990). Attitudes toward racial equality. *Humboldt Journal of Social Relations, 16*, 67-94.

The Center of Inquiry in the Liberal Arts at Wabash College. (2009). *Wabash National Study 2006-2009*. Retrieved Dec. 15th, 2010 from http://www.liberalarts.wabash.edu/study-overview/

Chamorro-Premuzic, T., & Furnham, A. (2003). Personality traits and academic examination performance. *European Journal of Personality, 17,* 237-250.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 9*(2), 233-255.

Chickering, A. W. (1969). *Education and identity* (1st ed.). San Francisco: Jossey-Bass.

Chickering, A. W., & Reisser, L. (1993). *Education and identity* (2nd ed.). San Francisco: Jossey-Bass.

Chun, M. (2002). Looking where the light is better: A review of the literature on assessing higher education quality. *Peer Review*, 4(2/3), 16-25.

Chun, M. (2006). *Faster, better, cheaper: The iron triangle of higher education assessment*. New York: The Council for Aid to Education.

Clark, B. R., & Trow, M. (1966). The organizational context. In T. M. Newcomb and E. K. Wilson (Eds.), *College peer groups: Problems and prospects for research* (pp. 17-70). Chicago: Aldine.

Clarke, M. (2002). News or noise? An analysis of U.S. News and World Report's ranking scores. *Educational Measurement: Issues and Practice, 21*(4), 39-48.

Colby, A., Ehrlich, T., Beaumont, E., & Stephens, J. (2003). *Educating citizens: Preparing America's undergraduates for lives of moral and civic responsibility*. San Francisco, CA: Jossey Bass.

College Board. (2008). *SAT-ACT Concordance Tables*. New York: Author. Retrieved July 3rd, 2010 from http://professionals.collegeboard.com/data-reports-research/sat/sat-act

College Portrait. (2009). *College Portrait of undergraduate education – About the College Portrait*. Retrieved May 20, 2010 from http://www.collegeportraits.org/about

Colquitt, J. A., LePine, J. A., & Noe, R. (2000). Toward an integrative theory of training motivation: A meta-analytic path analysis of 20 years of research. *Journal of Applied Psychology, 85,* 678-707.

Compton, M. (2012). *Everything you need to know about President Obama's Blueprint for College Affordability*. Washington, DC: The White House Blog. Retrieved January 28th, 2012 from http://www.whitehouse.gov/blog/2012/01/27/everything-you-need-know-about-president-obamas-blueprint-college-affordability

Cotic, M. & Zuljan, M. V. (2009). Problem-based instruction in mathematics and its impact on the cognitive results of the students and on affective-motivational aspects. *Educational Studies, 35*(3), 297-310.

Council for Aid to Education. (2009). *2008–2009 CLA Institutional Report: University College*. New York: Author. Retrieved April 22, 2011 from http://www.collegiatelearningassessment.org/files/CLA_0809_Institutional_Report_University_College.pdf

Council for Aid to Education. (2010). *2009–2010 CLA Institutional Report: University College*. New York: Author. Retrieved April 22, 2011 from http://collegiate learningassessment.org/files/CLA_0910_Report_University_College2.pdf

Council of Higher Education Accreditation. (2006). *CHEA at a glance*. Washington, DC: Author.

CPP, Inc. (2009). *CPP: Education*. Mountain View, CA: Author. Retrieved October 30, 2011 from https://www.cpp.com/en/education.aspx

Crystal, D., & DeBell, M. (2002). Sources of civic orientation among American youth: Trust, religion valuation, and attributions of responsibility. *Political Psychology, 23*(1), 113-132.

Cunha, F., & Heckman, J. J. (2010). Investing in our young people. In A. J. Reynolds, A. J. Rolnick, M. M. Englund, & J. A. Temple (Eds.), *Childhood programs and practices in the first decade of life: A human capital integration* (pp. 381-414). Cambridge, UK: Cambridge University Press.

DePascale, C. A. (2006). *Measuring growth with the MCAS tests: A consideration of vertical scales and standards.* Dover, NH: National Center for the Improvement of Educational Assessment.

Dewey, J. (1916). *Democracy and education: An introduction to the philosophy of education.* New York: The Free Press.

Dickeson, R. C. (2006). *The need for accreditation reform*. Retrieved Dec. 26th, 2010 from www2.ed.gov/about/bdscomm/list/hiedfuture/reports/dickeson.pdf

Dickeson, R. C. (2009). *Recalibrating the accreditation-federal relationship*. Retrieved Dec. 26th, 2010 from http://www.chea.org/pdf/2009_AC_Recalibrating_the_Accreditation-Federal_Relationship_Dickeson_wp.pdf

Duckworth, A., Seligman, M. (2005). Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychological Science, 16,* 939-944.

Dwyer, C. A., Millett, C. M., & Payne, D. G. (2006). *A culture of evidence: Postsecondary assessment and learning outcomes*. Princeton, NJ: Educational Testing Service.

Eaton, J. S. (2009). Accreditation in the United States. *New Directions for Higher Education, 145*, 79-86. DOI: 10.1002/he.

Efron, B. (1982). *The jackknife, bootstrap, and other resampling plans*. CBMS-NSF Regional conference series in applied mathematics, 38. SIAM, Philadelphia.

Efron, B. & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Monographs on statistics and applied probability, 57. New York: Chapman & Hall.

Ehrlich, T. (Ed.). (2000). *Civic responsibility and higher education*. Westport, CT: American Council on Education and Oryx Press.

Ekman, R., & Pelletier, S. (2008). Assessing student learning: A work in progress. *Change, 40*(4), 14-19.

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods, 12*(2), 121-138.

Ercikan, K. & Julian, M. (2002). Classification accuracy of assigning student performance to proficiency levels: Guidelines for assessment design. *Applied Measurement in Education, 15*(3), 269-294.

Erwin, T. D., & Wise, S. L. (2002). A scholar-practitioner model for assessment. In Banta, T. W. (Ed.), *Building a scholarship of assessment* (pp.67-81). San Francisco: Jossey-Bass.

Evans, N. J., Forney, D. S., Guido, F. M., Patton, L. D., & Renn, K. A. (2010). *Student development in college: Theory, research, and practice* (2nd ed.). San Francisco, CA: Jossey-Bass.

Evans, N. J., Forney, D. S., & Guido-DiBrito, F. (1998). *Student development in college: Theory, research, and practice*. San Francisco: Jossey-Bass.

Ewell, P., McClenney, K., & McCormick, A. C. (2011, September 20). Measuring engagement. *Inside Higher Ed*. Retrieved October 30, 2011 from www.insidehighered.com/views/2011/09/20/essay_defending_the_value_of_surveys_of_student_engagement

Fitzpatrick, L. (2009, August 20). Q&A: The man behind the U.S. News college rankings. *Times*. Retrieved April 14, 2011 from http://www.time.com/time/nation/article/0,8599,1917590,00.html

Fuller, W. A. (1987). *Measurement error models*. New York: John Wiley & Sons, Inc.

Garcia, P. (2007, November). *How to assess expected value-added: The CLA method*. Paper presented at the meeting of the California Association for Institutional Research, Monterey, CA.

Gong, B., Perie, M., & Dunn, J. (2006). *Using student longitudinal growth measures for school accountability under No Child Left Behind: An update to inform design decisions.* Dover, NH: National Center for the Improvement of Educational Assessment.

Gray, M. J., Ondaatje, E. H., Fricker, R., Geschwind, S., Goldman, C. A., Kaganoff, T., et al. (1999). *Combining service and learning in higher education: Evaluation of the learn and serve America, higher education program*. Santa Monica, CA: Rand Education.

Hagenaars, J. A., & McCutcheon, A. L. (Eds.). (2002). *Applied latent class analysis*. Cambridge, UK: Cambridge University Press.

Hambleton, R. K., & Slater, S. C. (1997). Reliability of credentialing examinations and the impact of scoring models and standard-setting policies. *Applied Measurement in Education, 10*, 19-39.

Hart Research Associates. (2010). *Raising the bar: Employers' views on college learning in the wake of the economic downturn*. Washington, DC: Author. Retrieved March 25, 2010 from www.aacu.org/leap/documents/2009_EmployerSurvey.pdf

Hearn, J. C. (1991). Academic and nonacademic influences on the college destinations of 1980 high school graduates. *Sociology of education, 64*(3), 158-171.

Heckman, J. J., & Rubinstein, Y. (2001). The importance of noncognitive skills: Lessons from the GED testing program. *The American Economic Review, 91*(2), 145-149.

Heider, F. (1958). *The psychology of interpersonal relations.* New York: Wiley.

Hersh, R. H., Bundick, M., Keeling, R., Keyes, C., Kurpius, A., Shavelson, R., et al. (2008). *A well-rounded education for a Flat World*. A paper prepared by the College Outcomes Project. Washington, DC: S. Engelhard Center. Retrieved March 25, 2010 from http://www.aacu.org/bringing_theory/documents/AWELLROUNDED-finalmasterFOOTNOTESINSERTEDnarrativecitationsA_Well-rounded-sept_25_Kurpiu_000.doc

Hershberg, T., Simon, V. A., & Lea-Kruger, B. (2004). The revelations of value-added: An assessment model that measures student growth in ways that NCLB fails to do. *The School Administrator, 61*(11), 10-14.

Higher Education Research Institute. (2011). *Higher Education Research Institute: 2012 College Senior Survey*. Retrieved May 20th, 2011 from http://www.heri.ucla.edu/cssoverview.php

Hoffman, D. M. (2009). Reflecting on social emotional learning: A critical perspective on trends in the United States. *Review of Educational Research, 79*(2), 533-556.

Hofmann, D. A., & Gavin, M. B. (1998). Centering decisions in hierarchical linear models: Implications for research in organizations. *Journal of Management, 24,* 623-641.

Hosch, B. J. (2010). *Time on test, student motivation, and performance on the Collegiate Learning Assessment: Implications for institutional accountability*. Paper presented at the Association for Institutional Research Annual Forum, Chicago, IL.

Hough, L. M. (2001). I/Owes its advances to personality. In B.W. Roberts & R. Hogan (Eds.), *Personality psychology in the workplace* (pp. 19-44).Washington, DC: American Psychological Association.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55.

Huang, Y. R., & Healy, C. C. (1997). The relation of Holland-typed majors to students' freshman and senior work values. *Research in Higher Education, 38*, 455-477.

Hurtado, S. (2007). Linking diversity with the educational and civic missions of higher education. *The Review of Higher Education, 30,* 185-196.

Hurtado, S., Newman, C. B., Tran, M. C., & Chang, M. J. (2010). Improving the rate of success for underrepresented racial minorities in STEM fields: Insights from a national project. *New Directions for Institutional Research, 148*, 5-15.

Hurtz, G. M., & Donovan, J. J. (2000). Personality and job performance: The big-five revisited. *Journal of Applied Psychology, 85,* 869-879.

Immerwahr, J. (2000). *Great expectations: How Californians view higher education*. San Jose, CA: National Center for Public Policy and Higher Education, and Public Agenda.

Indiana University. (2010). *Proposed resolution on adoption of AAC&U LEAP essential learning outcomes as the framework for general education programs on all campuses of Indiana University*. IN: Author. Retrieved March, 22nd, 2011 from http://www.iusb.edu/~sbas/senate_documents/03minutes/academic_senate_minutes/10-11/March2011documents/IU%20UFC%20Resolution%20on%20Gen Ed%20and%20LEAP%20-%2011.30.2010.pdf

Ingels, S., Curtin, T., Kaufman, P., Alt, M., & Chen, X. (2002). *Coming of age in the 1990s: The eighth-grade class of 1988 12 years later* (Statistical Analysis Report No. NCES 2002-321). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

Jacobi, M., Astin, A. W., & Ayala, F., Jr. (1987). *College student outcomes assessment: A talent development perspective*. ASHE-ERIC higher education report no. 7. Las Vegas: Association for the Study of Higher Education.

Jones, D. A. (2009). Higher education assessment – Who are we assessing, and for what purpose? *Peer Review, 11*(1), 35.

Jöreskog, K. G. (2005). *Structural equation modeling with ordinal variables using LISREL*. Lincolnwood, IL: Scientific Software International, Inc. Retrieved June 20, 2011 from http://www.ssicentral.com/lisrel/ordinal.htm

Jöreskog, K. G. & Sörbom, D. (2001). *LISREL 8: User's reference guide*. Lincolnwood, IL: Scientific Software International, Inc.

Jöreskog, K. G. & Sörbom, D. (2006). *LISREL 8.80 for Windows [Computer Software]*. Lincolnwood, IL: Scientific Software International, Inc.

Keeter, S., Zukin, C., Andolina, M., & Jenkins, K. (2002). *The civic and political health of the nation: A generational portrait.* Medford, MA: The Center for Information & Research on Civic Learning & Engagement.

Kegan, R. (1994). *In over our heads: The mental demands of modern life.* Cambridge, MA: Harvard University Press.

Kelley, H. H. (1973). The processes of causal attribution. *American Psychologist, 28*, 107-128.

Kelly, G. A. (1955). *The psychology of personal constructs* (Vol.1). New York: Norton.

King, P. M., Kendall Brown, M., Lindsay, N. K., & Vanhecke, J. R. (2007). Liberal arts student learning outcomes: An integrated approach. *About Campus, 12*(4), 2-9.

King, P. M., & Kitchener, K. S. (1994). *Developing reflective judgment: Understanding and promoting intellectual growth and critical thinking in adolescents and adults*. San Francisco, CA: Jossey-Bass.

Klein, S. (2002). Direct assessment of cumulative student learning. *Peer Review, 4*, 26-28.

Klein, S. (2009). *The Lumina longitudinal study: Summary of procedures and findings*. Retrieved April 10th, 2010 from http://www.collegiatelearningassessment.org/files/CLA_Lumina_Longitudinal_Study_Summary_Findings.pdf

Knox, W. E., Lindsay, P., & Kolb, M. N. (1993). *Does college make a difference? Long-term changes in activities and attitudes*. Westport, CT: Greenwood Press.

Kobrin, J. L., Patterson, B. F., Shaw, E. J., Mattern, K. D., & Barbuti, S. M. (2008). *Validity of the SAT for predicting first-year college grade point average*. College Board Research Report No. 2008-5. New York: College Board.

Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In D. A. Goslin (Ed.), *Handbook of socialization theory and research* (pp. 347-480). Chicago: Rand McNally.

Kolen, M. J., Hanson, B. A., & Brennan, R. L. (1992). Conditional standard errors of measurement for scale scores. *Journal of Educational Measurement, 29*, 285-307.

Kreft, I. G. G., de Leeuw, J., & Aiken, L. S. (1995). The effect of different forms of centering in hierarchical linear models. *Multivariate Behavioral Research, 30,* 1-21.

Kuh, G. (2006). Engagement: The bridge from here to there. Director's message in *Engaged learning: fostering success for all students*. Bloomington, IN: National Survey of Student Engagement.

Kuh, G. D., Hayek, J. C., Carini, R. M., Ouimet, J. A., Gonyea, R. M., & Kennedy, J. (2001). *NSSE technical and norms report*. Bloomington, IN: Indiana University Center for Postsecondary Research and Planning.

Kuh, G. D., Hu, S., & Vesper, N. (2000). "They shall be known by what they do": An activities-based typology of college students. *Journal of College Student Development, 41*, 228-244.

Kuh, G. D., & Ikenberry, S. (2009). *More than you think, less than we need: Learning outcomes assessment in American higher education*. Urbana-Champaign, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment.

Kyllonen, P. C. (2005). The case for noncognitive assessments. *ETS R&D Connections*. Princeton, NJ: Educational Testing Service.

Kyllonen, P. C., Walters, A. M., & Kaufman, J. C. (2005). Noncognitive constructs and their assessment in graduate education: A review. *Educational Assessment, 10*(3), 153-184.

LaNasa, S. M., Olson, E., & Alleman, N. (2007). The impact of on-campus student growth on first-year student engagement and success. *Research in Higher Education, 48*(8), 941-966.

Larson, G. (2008). *Assessing the CLA: An evaluation of the psychometric properties of the Collegiate Learning Assessment's "value-added" in cross sectional administrations*. Paper presented at the 2008 Association for Institutional Research Forum, Seattle, WA.

Lawlor, D. A., Clark, H., Ronalds, G., & Leon, D. A. (2006). Season of birth and childhood intelligence: Findings from the Aberdeen children of the 1950s cohort study. *British Journal of Educational Psychology, 76*(3), 481-499.

Lederman, D., & Redden, E. (2007, January 31). Accountability and comparability. *Inside Higher Ed.* Retrieved Dec. 26th, 2010 from http://www.insidehighered.com/news/2007/01/31/compare

Leef, G. C., & Burris, R. D. (2002). *Can college accreditation live up to its promise?* Washington, DC: American Council of Trustees and Alumni.

Lenning, O. T. (1977). *The outcomes structure: An overview and procedures for applying it in postsecondary education institutions*. Boulder, CO: National Center for Higher Education Management Systems.

Levin, H. M. (2011). *The utility and need for incorporating non-cognitive skills into large scale educational assessments*. Paper presented at the International Large-Scale Assessment Conference, Princeton, NJ.

Lewis, H. R. (1997). *The mission of Harvard College*. Retrieved March 22nd, 2011 from http://www.harvard.edu/siteguide/faqs/faq110.php

Linacre, J. M. (2006). *WINSTEPS Rasch measurement computer program (Version 3.64.2)*. Chicago: Winsteps.com.

Linacre, J. M. (2011). *A user's guide to WINSTEPS /MINISTEP Rasch-model computer programs*. Chicago: *Winsteps*.com.

Lissitz, R. W., Doran, H., Schafer, W. D., & Wilhoft, J. (2006). Growth modeling, value added modeling, and linking: An introduction. In Lissitz, R. W. (Ed.), *Longitudinal and value-added models of student performance* (pp. 1-46). Maple Grove, MN: JAM Press.

Little, R. J. A. (1993). Post-stratification: A modeler's perspective. *Journal of the American Statistical Association, 88*, 1001-1012.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical Analysis with Missing Data* (2nd Ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Liu, O. L. (2008). *Measuring learning outcomes in higher education using the Measure of Academic Proficiency and Progress (MAPP™).* ETS Research Report No. RR-08-047. Princeton, NJ: Educational Testing Service.

Liu, O. L. (2011). *Examining American post-secondary education*. ETS Research Report No. RR-11-22. Princeton, NJ: Educational Testing Service.

Liu, O. L., Minsky, J., Ling, G., & Kyllonen, P. (2009). Using the standardized letters of recommendation in selection. *Educational and Psychological Measurement, 69*(3), 475-492.

Lopez, M. H., & Brown, B. (2006). *Civic engagement among 2-year and 4-year college students*. College Park, MD: The Center for Information & Research on Civic Learning & Engagement.

Lopez, M. H., & Elrod, B. A. (2005). *College attendance and civic engagement among 18-25 year olds*. College Park, MD: The Center for Information & Research on Civic Learning & Engagement.

Lopez, M. H., & Elrod, B. A. (2006). *College attendance and civic engagement among 18 to 25 year olds*. College Park, MD: The Center for Information & Research on Civic Learning & Engagement.

Luo, J., & Jamieson-Drake, D. (2005, November 3). Linking student precollege characteristics to college development outcomes: The search for a meaningful way to inform institutional practice and policy. *IR Applications, 7*, Association for Institutional Research.

Machung, A. (1998). Playing the rankings game. *Change, 30*(4), 12-16,

Malandra, G. H. (2008). Accountability and learning assessment in the future of higher education. *On The Horizon, 16* (2), 57-71. DOI: 10.1108/10748120810874478.

Marcon, R. A. (1999). Differential impact of preschool models on development and early learning of inner-city children: A three-cohort study. *Developmental Psychology, 35*(2), 358 -375.

Martin, M. O., & Preuschoff, C. (2008). Creating the TIMSS 2007 background indices. In J. F Olson, M. O. Martin, & I. V. S. Mullis (Eds.), *TIMSS 2007 Technical Report* (pp. 281-338). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Martin, W. E., Jr., Swartz-Kulstad, J. L., & Madson, M. (1999). Psychosocial factors that predict the college adjustment of first-year undergraduate students: Implications for college counselors. *Journal of College Counseling, 2*(2), 121-133.

Masters, G. N., & Wright, B. D. (1997). The partial credit model. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 101-121). New York: Springer-Verlag.

McCaffrey, D. F., Lockwood, J. R., Koretz, D. M., & Hamilton, L. S. (2003). *Evaluating value-added models for teacher accountability*. Santa Monica: RAND, MG-158-EDU.

McDaniel, M. A., Hartman, N. S., Whetzel, D. L., & Grubb, W. L., III (2007). Situational judgment tests, response instructions, and validity: A meta-analysis. *Personnel Psychology, 60*(1), 63-91.

McPherson, P., & Shulenburger, D. (2006). *Elements of accountability for public universities and colleges*. National Association of State Universities and Land-Grant Colleges. Retrieved Dec. 16th, 2010 from http://www.voluntarysystem.org/docs/background/DiscussionPaper2_July06.pdf

Menard, S. (1991). *Longitudinal research*. Newbury Park, CA: Sage.

Mentkowski, M., & Associates. (2000). *Learning that lasts: Integrating learning, development, and performance in college and beyond.* San Francisco: Jossey-Bass.

Mentkowski, M., & Doherty, A. (1984). *Careering after college: Establishing the validity of abilities learned in college for later careering and professional performance*. Final report to the national institute of education. ED252144.

Messick, S. (1979). Potential uses of noncognitive measurement in education. *Journal of Educational Psychology, 71*(3), 281-292.

Miller-Bernal, L. (1993). Single-sex versus coeducational environments: A comparison of women students' experiences at four colleges. *American Journal of Education, 102*, 23-54.

Millett, C. M., Stickler, L. M., Payne, D. G., & Dwyer, C. A. (2007). *A culture of evidence: Critical features of assessment for postsecondary student learning.* Princeton, NJ: Educational Testing Service.

Millman, J. (Ed.) (1997). *Grading teachers*, *grading schools: Is student achievement a valid evaluation measure?* Thousand Oaks, CA: Corwin Press.

Monks, J., & Ehrenberg, R. G. (1999). *The impact of US. News and World Report college rankings on admissions outcomes and pricing policies at selective private institutions*. Working Paper No. 7227. Cambridge, MA: National Bureau of Economic Research.

Morse, R. (2010, August 17). Methodology: Undergraduate ranking criteria and weights. *U.S. News and World Report*. Retrieved April 6th, 2011 from http://www.usnews.com/education/articles/2010/08/17/methodology-undergraduate-ranking-criteria-and-weights-2011

Morse, S. (1989). *Renewing civic capacity: Preparing college students for service and citizenship*. ASHE-ERIC Higher Education Report No. 8. Washington, DC: The University of George Washington.

Myers-Lipton, S. J. (1998). Effect of a comprehensive service-learning program on students' civic responsibility. *Teaching Sociology, 2*, 243-258.

NASPA & ACPA. (2004). *Learning reconsidered: A campus-wide focus on the student experience*. Washington, DC: The National Association of Student Personnel Administrators, The American College Personnel Association.

National Center for Education Statistics. (2011). *The Integrated Postsecondary Education Data System: About IPEDS*. Retrieved May 20th, 2011 from http://nces.ed.gov/ipeds/about

National Opinion Research Center. (1997). *A review of the methodology for the U.S. News & World Report's rankings of undergraduate colleges and universities*. Chicago: Author.

National Survey of Student Engagement. (2004). *Student engagement: Pathways to collegiate success – 2004 annual survey results*. Bloomington, IN: Indiana University Center for Postsecondary Research.

National Survey of Student Engagement. (2008). *Promoting engagement for all students: The imperative to look within: 2008 results.* Bloomington, IN: Indiana University Center for Postsecondary Research.

National Survey of Student Engagement. (2011). *NSSE's commitment to data quality*. Bloomington, IN: Indiana University Center for Postsecondary Research. Retrieved October 30th, 2011 from http://nsse.iub.edu/_/?cid=154

Neihart, M. (2007). The socioaffective impact of acceleration and ability grouping: Recommendations for best practice. *Gifted Child Quarterly, 51*(4), 330-341.

Nelson Laird, T. F., Engberg, M. E., & Hurtado, S. (2005). Modeling accentuation effects: Enrolling in a diversity course and the importance of social action engagement. *The Journal of Higher Education, 76*(4), 448-476.

Nettles, M. T., Cole, J. K., & Sharp, S. (1997). *Assessment of teaching and learning in higher education and public accountability: State governing, coordinating board & regional accreditation association policies and practices*. Stanford, CA: Stanford University, National Center for Postsecondary Improvement.

Newman, F. (1985). *Higher education and the American resurgence*. A Carnegie Foundation Special Report. Princeton, NJ: The Carnegie Foundation for the Advancement of Teaching.

No Child Left Behind Act of 2001. (2002). Pub. L. No. 107-110, 115 Stat. 1425, 20 U.S.C. § 6301 et seq.

Noble, J., & Sawyer, R. (2002). *Predicting different levels of academic success in college using high school GPA and ACT Composite score.* ACT Research Report 2002-4. Iowa City, IA: ACT.

Nolin, M. J., Chaney, B., & Chapman, C. (1997). *Student participation in community service activity*. NCES 97-331. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Nora, A. (Ed.). (2011). *The Review of Higher Education [Special issue], 35*(1).

Obama, B. (2012, January). *Remarks by the President on college affordability*. Speech presented at the University of Michigan, Ann Arbor, MI. Retrieved January 28th, 2012 from http://www.whitehouse.gov/the-press-office/2012/01/27/remarks-president-college-affordability-ann-arbor-michigan

Obler, S. S., Slark, J., & Umbdenstock, L. (1993). Classroom assessment. In T. W. Banta, and Associates. (Eds.). *Making a difference: Outcomes of a decade of assessment in higher education* (pp. 211-226). San Francisco: Jossey-Bass.

O'Dwyer, L. M., Carey, R., & Kleiman, G. (2007). A study of the effectiveness of the Louisiana Algebra I online course. *Journal of Research on Technology in Education*, *39*(3), 289-306.

O'Neil, H. F., Abedi, J., Miyoshi, J., & Mastergeorge, A. (2005). Monetary incentives for low-stakes tests. *Educational Assessment, 10*(3), 185-208.

Organization for Economic Cooperation and Development. (2008a). *Measuring improvements in learning outcomes: Best practices to assess the value-added of schools*. Paris: OECD Publications.

Organization for Economic Cooperation and Development. (2008b). *PISA 2006 technical report*. Paris: OECD Publications.

Organization for Economic Cooperation and Development. (2010-2011). *OECD Assessment of Higher Education Learning Outcomes brochure*. Retrieved August 16th, 2011 from http://www.oecd.org/dataoecd/37/49/45755875.pdf

Organization for Economic Cooperation and Development. (2011). *Testing student and university performance globally: OECD's AHELO*. Retrieved August 16th, 2011 from http://www.oecd.org/edu/ahelo

Oswald, F. L., Schmitt, N., Kim, B. H., Ramsay, L. J., & Gillespie, M. A. (2004). Developing a biodata measure and situational judgment inventory as predictors of college student performance. *Journal of Applied Psychology, 89*, 187-207.

Outcalt, C. L., & Skewes-Cox, T. E. (2002). Involvement, interaction, and satisfaction: The human environment at HBCUs. *The Review of Higher Education, 25*(3), 331-347.

Pascarella, E. T. (2006). How college affects students: Ten directions for future research. *Journal of College Student Development, 47*(5), 508-520.

Pascarella, E. T., Ethington, C. A., & Smart, J. C. (1988). The influence of college on humanitarian/civic involvement values. *The Journal of Higher Education, 59*(4), 412-437.

Pascarella, E. T., Seifert, T. A., & Blaich, C. (2009). How effective are the NSSE benchmarks in predicting important educational outcomes? *Change, 42*(1), 16-22.

Pascarella, E. T., & Terenzini, P. T. (1991). *How college affects students: Findings and insights from twenty years of research*. San Francisco: Jossey-Bass.

Pascarella, E. T., & Terenzini, P. T. (2005). *How college affects students: A third decade of research*. San Francisco: Jossey-Bass.

Pascarella, E. T., Wolniak, G. C., Seifert, T. A., Cruce, T. M., & Blaich, C. F. (2005*). Liberal arts colleges and liberal arts education: New evidence on impacts*. ASHE higher education report, 31(3). San Francisco: Jossey-Bass.

Perry, W. G. (1970). *Forms of intellectual and ethical development in the college years: A scheme*. New York: Holt, Rinehart and Winston.

Perry, W. G. (1981). Cognitive and ethical growth: The making of meaning. In A. W. Chickering & Associates (Eds.), *The modern American college* (pp. 76-116). San Francisco: Jossey-Bass.

Pew Partnership for Civic Change. (2004). *New directions in civic engagement: University avenue meets main street*. Charlottesville, VA: Author.

Pike, G. R. (2006). Assessment measures: Value-added models and the Collegiate Learning Assessment. *Assessment Update, 18*(4), 5-7.

Pike, G. R., Kuh, G. D., & McCormick, A. C. (2011). An investigation of the contingent relationships between learning community participation and student engagement. *Research in Higher Education, 52*(3), 300 -322.

Preuschoff, A. C. (2010). Using TIMSS and PIRLS to construct global indicators of effective environments for learning (Doctoral dissertation, Boston College, 2010). *Dissertation Abstracts International, 72*(02), 315.

PROC LCA & PROC LTA (Version 1.2.7) [Software]. (2011). University Park: The Methodology Center, Penn State. Retrieved May 7, 2011 from http://methodology.psu.edu

Pryor, J. H., Hurtado, S., DeAngelo, L., Blake, L. P., & Tran, S. (2010). *The American freshman: National norms fall 2010*. Los Angeles: Higher Education Research Institute, UCLA.

Putnam, R. D. (2000) *Bowling alone: The collapse and revival of American community*. New York: Simon and Schuster.

Ratcliff, J. L., Jones, E. A., Guthrie, D. S., & Oehler, D. (1991). *The effect of coursework patterns, advisement, and course selection on the development of general learned abilities of college graduates*. University Park, PA: The Pennsylvania State University, National Center on Postsecondary Teaching, Learning, and Assessment.

Raudenbush, S. W. (1989). "Centering" predictors in multilevel analysis: Choices and consequences. *Multilevel Modelling Newsletter, 1*(2), 10-12.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods.* Thousand Oaks, CA: Sage publications.

Raudenbush, S. W., Bryk, A. S, Cheong, Y. F., Congdon, R., & du Toit, M. (2004). *HLM 6: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International, Inc.

Raudenbush, S. W., Bryk, A. S, & Congdon, R. (2004). *HLM 6 for Windows* [Computer software]. Lincolnwood, IL: Scientific Software International, Inc.

Reuben, J. A. (1996). *The making of the modern university: Intellectual transformation and the marginalization of morality*. Chicago: The University of Chicago Press.

Rhee, B., & Dey, E. L. (1996, November). *Collegiate influences on the civic values of students*. Paper presented at the meeting of the Association for the Study of Higher Education. Memphis, TN.

Robbins, S. B., Allen, J., Casillas, A., Peterson, C. H., & Le, H. (2006). Unraveling the differential effects of motivational and skills, social, and self-management measures from traditional predictors of college outcomes. *Journal of Educational Psychology, 98*(3), 598-616.

Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin, 130,* 261-288.

Roberts, B. W., & Hogan, R. (Eds.). (2001). *Personality psychology in the workplace. Decade of behavior*. Washington, DC: American Psychological Association.

Roberts, J. K. (2007, April). *Group dependency in the presence of small intraclass correlation coefficients: An argument in favor of NOT interpreting the ICC*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Robinson, W. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review, 15,* 351-357.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: John Wiley & Sons, Inc.

Salgado, J. F. (1998). Big five personality dimensions and job performance in army and civil occupations: A European perspective. *Human Performance, 11*, 271-288.

Sanders, W. L., Saxton, A. M., & Horn, S. P. (1997). The Tennessee Value-Added Assessment System (TVAAS): A quantitative, outcomes-based approach to educational assessment. In Millman, J. (Ed.), *Grading teachers, grading schools: Is student achievement a valid evaluation measure?* (pp. 137-162). Thousand Oaks, CA: Corwin Press.

SAS. (n.d.). *Schooling Effectiveness, SAS® EVAAS® for K-12.* Retrieved August 17, 2011 from http://www.sas.com/govedu/edu/k12/evaas/index.html

Sax, L. J. (2000). Citizenship development and the American college student. In T. Ehrlich (Ed.), *Civic responsibility and higher education* (pp. 3-18). Westport, CT: American Council on Education and Oryx Press.

Sax, L. J., & Astin, A. W. (1997). The benefits of service: Evidence from undergraduates. *Educational Record, 78*, 25-32.

Sax, L. J., & Astin, A. W. (1998). Developing "civic virtue" among college students. In J. Gardner & G. Van der Veer (Eds.), *The Senior Year Experience* (pp. 133-151). San Francisco: Jossey-Bass.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data* (Monographs on Statistics and Applied Probability, 72). Boca Raton, FL: Chapman & Hall/CRC.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*(2), 262-274.

Schulz, W., Ainley, J., Fraillon, J., Kerr, D. & Losito, B. (2010). *Initial Findings from the IEA International Civic and Citizenship Education Study.* Amsterdam: IEA.

Schulz, W., & Brese, F. (2008, March). *Assessing student knowledge, background and perceptions in the International Civic and Citizenship Study*. Paper prepared at the annual meeting of the American Educational Association, New York, NY.

Sedlacek, W. E. (2004). *Beyond the big test: Noncognitive assessment in higher education*. San Francisco: Jossey-Bass.

Sedlacek, W. E. (2005). The case for noncognitive measures. In W. Camara & E. Kimmel (Eds.), *Choosing students: Higher education admission tools for the 21st century* (pp. 177-193). Mahwah, NJ: Lawrence Erlbaum.

Seifert, T. A., Goodman, K., King, P. M., & Baxter Magolda, M. (2010). Using mixed methods to study first-year college impact on liberal arts learning outcomes. *Journal of Mixed Methods Research, 4*(3), 248-267.

Seifert, T. A., Pascarella, E. T., Erkel, S. I., & Goodman, K. M. (2010, Winter). The importance of longitudinal pretest-posttest designs in estimating college impact. *New directions in institutional research: Longitudinal assessment for institutional improvement,* s2, 5-16.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton-Mifflin.

Sharkness, J., DeAngelo, L., & Pryor, J. (2010). *CIRP construct technical report*. University of California, Los Angeles: Higher Education Research Institute. Retrieved April 15th, 2010 from http://www.heri.ucla.edu/PDFs/constructs/technicalreport.pdf and http://www.heri.ucla.edu/PDFs/constructs/2009%20Appendix.pdf

Shavelson, R. J. (2010). *Measuring college learning responsibly: Accountability in a new era*. Stanford, CA: Stanford University Press.

Shermis, M. D. (2008). The Collegiate Learning Assessment: A critical perspective. *Assessment Update, 20*(2), 10-12.

Shulenburger, D., & Keller, C. (2009). *Interpretation of findings of the Test Validity Study conducted for the Voluntary System of Accountability*. Washington, DC: Voluntary System of Accountability. Retrieved August 19th, 2011 from www.voluntarysystem.org/docs/reports/VSAabstract_TVS.pdf

Slotnik, D. E., & Pérez-peña, R. (2012, January 31). College says it exaggerated SAT figures for ratings. *The New York Times*, A12.

Smith, D. G., Wolf, L. E., & Morrison, D. E. (1995). Paths to success: Factors related to the impact of women's colleges. *Journal of Higher Education, 66*, 245-266.

Smith, M. B., Nowacek, R. S., & Bernstein, J. L. (Eds.). (2010). *Citizenship across the curriculum*. Bloomington, IN: Indiana University Press.

Steedle, J. T. (2009). *Advancing institutional value-added score estimation*. New York: Council for Aid to Education.

Steedle, J. T. (2010, April). *Incentives, motivation, and performance on a low-stakes test of college learning*. Presented at the 2010 annual meeting of the American Educational Research Association, Denver, CO.

Sundre, D. L., & Moore, D. L. (2002). The student opinion scale: A measure of examinee motivation. *Assessment Update, 14*(1), 8-9.

Szelényi, K., Bryant, A. N., & Lindholm, J. A. (2005). What money can buy: Examining the effects of prepaid monetary incentives on survey response rates among college students. *Educational Research and Evaluation, 11*(4), 385-404.

Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology, 44,* 703-742.

Thompson, N. (2000, September). Playing with numbers: How U.S. News and World Report mismeasures higher education and the rest of us. *Washington Monthly*. Retrieved April 14, 2011 from http://www.washingtonmonthly.com/features /2000/0009.thompson.html

Thomson, G., & Douglass, J. A. (2009). *Decoding learning gains: Measuring outcomes and the pivotal role of the major and student backgrounds*. Research & Occasional Paper Series: CSHE.5.09. Center for Studies in Higher Education. Berkeley, CA: University of California, Berkeley.

Tong, Y., & Kolen, M. J. (2007). Comparisons of methodologies and results in vertical scaling for educational achievement tests. *Applied Measurement in Education*, *20*(2), 227-253.

Torney-Purta, J., & Amadeo, J. (2011, March). *The contributions of international large-scale studies in civic engagement and citizenship*. Paper presented at the International Large-Scale Assessment Conference, Princeton, NJ.

Tracey, T. J., & Sedlacek, W. E. (1984). Noncognitive variables in predicting academic success by race. *Measurement and Evaluation in Guidance*, 16, 171-178.

Tracey, T. J., & Sedlacek, W. E. (1988). A comparison of White and Black student academic success using noncognitive variables: A LISREL analysis. *Research in Higher Education. 27*, 333-348.

Truxillo, C. (2005). *Maximum likelihood parameter estimation with incomplete data*. SAS Users Group International Conference, Philadelphia, PA. Available at http://www2.sas.com/proceedings/sugi30/111-30.pdf

Tukey, J. W. (1953). The problem of multiple comparisons. Unpublished manuscript. In *The Collected Works of John W. Tukey VIII. Multiple Comparisons*: *1948–1983*, pp.1-300. New York: Chapman & Hall.

University of Michigan. (2011). *Mission statement & vision statement*. MI: Author. Retrieved March 22nd, 2011 from http://www.umich.edu/pres/mission.php

University of Wisconsin System – Board of Regents. (2010). *Mission statements of UW System campuses and institutions*. WI: Author. Retrieved March 22nd, 2011 from http://www.wisconsin.edu/ about/mission.htm#univcluster

U.S. Department of Education. National Center for Education Statistics. (1996). *The national assessment of college student learning: An inventory of state-level assessment activities. A report of the proceedings of the third study design workshop*. NCES 96-862, by Sal Corrallo, project director. Washington, DC: Author.

U.S. Department of Education. National Center for Education Statistics. (1997). *Student outcome information for policy-making*. NCES 97-991. Prepared by P. Terenzini for the Council of the National Postsecondary Education Cooperative Working Group on Student Outcomes from a Policy Perspective. Washington, DC: Author.

U.S. Department of Education. (2006). *A test of leadership: Charting the future of higher education.* A Report of the Commission Appointed by Secretary of Education Margaret Spellings. Washington, DC: Author.

U.S. Department of Education. (2010). *Race to the Top assessment program executive summary*. Washington, DC: Author.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*, 4-69.

Vermunt, J. K., & Magidson, J. (2002). Latent class cluster analysis. In J. A. Hagenaars & A. L. McCutcheon (Eds.), *Applied latent class analysis* (pp.89-106). Cambridge, UK: Cambridge University Press.

Villalpando, O. (1996). *The long term effects of college on Chicana and Chicano students' "other oriented" values, service careers and community involvement* (Doctoral dissertation, University of California, Los Angeles, 1996). *Dissertation Abstracts International, 57*(06), 256.

Voluntary System of Accountability. (2008a). *Voluntary System of Accountability (VSA): Information on learning outcomes measures*. Retrieved March 20, 2010, from http://www.voluntarysystem.org/docs/cp/LearningOutcomesInfo.pdf

Voluntary System of Accountability. (2008b). *The Voluntary System of Accountability (VSA^{SM}): Overview of College Portrait*. Retrieved March 20, 2011, from http://www.voluntarysystem.org/docs/cp/CollegePortraitOverview.pdf

Voluntary System of Accountability. (2010). *VSA participation agreement*. Retrieved August 16, 2011, from http://www.voluntarysystem.org/content/VSAParticipationAgreement.htm

Voluntary System of Accountability. (2011). *Voluntary System of Accountability Program®: Undergraduate education reports – VSA participants by state*. Retrieved August 16, 2011, from http://www.voluntarysystem.org/index.cfm?page=templates

Voluntary System of Accountability. (n.d.). *Voluntary System of Accountability Program®: Undergraduate education reports – The VSA online*. Retrieved March 20, 2010, from http://www.voluntarysystem.org/index.cfm

Wagerman, S. A., & Funder, D. C. (2007). Acquaintance reports of personality and academic achievement: A case for conscientiousness. *Journal of Research in Personality, 41*, 221-229.

Wainer, H. (Ed.). (2004). Value Added Assessment Special Issue. *Journal of Educational and Behavioral Statistics, 29*(1).

Wall-Smith, S. B. (2006). The Collegiate Learning Assessment at Harris-Stowe State University. *Assessment Update, 18*(3), 14-15.

Walpole, M. B., Burton, N. W., Kanyi, K., & Jackenthal, A. (2002). *Selecting successful graduate students: In-depth interviews with GRE users.* Graduate Record Examination Board Research Report No. 99-11R. Educational Testing Service Research Report 02-08. Princeton, NJ: Educational Testing Service.

Wang, Y. (2009, April). *The effect of test difficulty and reliability on detecting student growth*. Poster presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Webster, W. & Mendro, R. (1997). The Dallas value-added accountability system. In J. Millman (Ed.), *Grading teachers, grading schools: Is student achievement a valid evaluation measure?* (pp. 81-99). Thousand Oaks, CA: Corwin Press, Inc.

Weiner, B. (1972). Attribution theory, achievement motivation, and the educational process. *Review of Educational Research, 42,* 203-215.

Western Association of Schools and Colleges. (2008). *Handbook of Accreditation*. Alameda, CA: Author.

Willingham, W. W. (1985). *Success in college: The role of personal qualities and academic ability.* New York: College Entrance Examination Board.

Wilson, K. L., Rest, J. R., Boldizar, J. P., & Deemer, D. K. (1992). Moral judgment development: The effects of education and occupation. *Social Justice Research, 5*(1), 31-48.

Wise, S. L., & DeMars, C. E. (2005). Low examinee effort in low-stakes assessment: Problems and potential solutions. *Educational Assessment, 10*(1), 1-17.

Wolf, L. F., & Smith, J. K. (1995). The consequence of consequence: Motivation, anxiety, and test performance. *Applied Measurement in Education, 8*, 227-242.

Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions, 8*(3), 370.

Yale University. (1976). *The Yale corporation: Charter and legislation*. New Haven: Author.

Zhang, L. (1999). A note on post-stratification when analyzing binary survey data subject to nonresponse. *Journal of Official Statistics, 15*(2), 329-334.

APPENDIX A. THE CIRP SURVEY INSTRUMENTS

PLEASE PRINT NAME AND PERMANENT/HOME ADDRESS (one letter or number per box)

NAME: FIRST | M I | LAST | When were you born? Month (01-12) Day (01-31) Year

ADDRESS: STATE: ZIP: PHONE: – –

2002 STUDENT INFORMATION FORM

MARKING DIRECTIONS

Your responses will be read by an optical mark reader. Please,

- Use a pencil or black or blue pen.
- Fill in the oval completely.
- Erase cleanly any marks you wish to change or "X" out mark if in pen.

CORRECT MARK ● INCORRECT MARKS ✓ X

Dear Student:

This information is being collected as part of a continuing study of higher education conducted by the American Council on Education and the University of California at Los Angeles. Your participation in this research will help us to achieve a better understanding of how students are affected by their college experiences. Detailed information on this research program is available from the Higher Education Research Institute at UCLA. Identifying information has been requested in order to make subsequent mail follow-up studies possible. Your responses are held in the strictest professional confidence.

Sincerely,

Alexander W. Astin, Director
Higher Education Research Institute

PLEASE USE PENCIL OR BALLPOINT PEN

PLEASE PROVIDE YOUR SOCIAL SECURITY NO.	Mark here if directed	
	GROUP CODE A	GROUP CODE B

1. Your sex: ○ Male ○ Female

2. How old will you be on December 31 of this year? (Mark one)

16 or younger ○ | 17 ○ | 18 ○ | 19 ○ | 20 ○ | 21-24 ○ | 25-29 ○ | 30-39 ○ | 40-54 ○ | 55 or older ○

3. Is English your native language?

○ Yes ○ No

4. In what year did you graduate from high school? (Mark one)

2002 ○ | 2001 ○ | 2000 ○ | 1999 or earlier ○ | Did not graduate but passed G.E.D. test ○ | Never completed high school ○

5. Are you enrolled (or enrolling) as a: (Mark one)

Full-time student? ○ | Part-time student? ○

6. How many miles is this college from your permanent home? (Mark one)

5 or less ○ | 6-10 ○ | 11-50 ○ | 51-100 ○ | 101-500 ○ | Over 500 ○

7. What was your average grade in high school? (Mark one)

A or A+ ○ | A– ○ | B+ ○ | B ○ | B– ○ | C+ ○ | C ○ | D ○

8. What were your scores on the SAT I and/or ACT?

SAT VERBAL | SAT MATH | ACT Composite

9. Citizenship status:

○ U.S. citizen
○ Permanent resident (green card)
○ Neither

10. Have you had, or do you feel you will need, any special tutoring or remedial work in any of the following subjects? (Mark all that apply)

	Have Had	Will Need
English	○	○
Reading	○	○
Mathematics	○	○
Social Studies	○	○
Science	○	○
Foreign Language	○	○
Writing	○	○

11. Prior to this term, have you ever taken courses for credit at this institution?

○ Yes ○ No

12. Since leaving high school, have you ever taken courses at any other institution? (Mark all that apply in each column)

	For Credit	Not for Credit
Yes, at a community/junior college	○	○
Yes, at a 4-yr. college or university	○	○
Yes, at some other postsecondary school (For example, technical, vocational, business)	○	○

13. Where do you plan to live during the fall term? (Mark one)

With my family or other relatives ○
Other private home, apartment or room ○
College dormitory ○
Fraternity or sorority house ○
Other campus student housing ○
Other ○

14. Is this college your: (Mark one)

First choice? ○ | Second choice? ○ | Third choice? ○ | Less than third choice? ○

15. To how many colleges other than this one did you apply for admission this year?

None ○ | 1 ○ | 2 ○ | 3 ○ | 4 ○ | 5 ○ | 6 ○ | 7-10 ○ | 11 or more ○

16. Do you have a disability? (Mark all that apply)

None ○
Hearing ○
Speech ○
Orthopedic ○
Learning disability ○
Health-related ○
Partially sighted or blind ○
Other ○

17. Did your high school require community service for graduation?

○ Yes ○ No

18. What is the highest academic degree that you intend to obtain? (Mark one in each column)

	Highest Planned	Highest Planned at this college
None	○	○
Vocational certificate	○	○
Associate (A.A. or equivalent)	○	○
Bachelor's degree (B.A., B.S., etc.)	○	○
Master's degree (M.A., M.S., etc.)	○	○
Ph.D. or Ed.D.	○	○
M.D., D.O., D.D.S., or D.V.M.	○	○
LL.B. or J.D. (Law)	○	○
B.D. or M.DIV. (Divinity)	○	○
Other	○	○

19. Are your parents: (Mark one)

Both alive and living with each other? ○
Both alive, divorced or living apart? ○
One or both deceased? ○

20. How much of your first year's educational expenses (room, board, tuition, and fees) do you expect to cover from each of the sources listed below? (Mark one answer for each possible source)

	None	Less than $1,000	$1,000-2,999	$3,000-5,999	$6,000-9,999	$10,000+
Family resources (parents, relatives, spouse, etc.)	○	○	○	○	○	○
My own resources (savings from work, work-study, other income)	○	○	○	○	○	○
Aid which need not be repaid (grants, scholarships, military funding, etc.)	○	○	○	○	○	○
Aid which must be repaid (loans, etc.)	○	○	○	○	○	○
Other than above	○	○	○	○	○	○

21. What is your best estimate of your parents' total income last year? Consider income from all sources before taxes. (Mark one)

○ Less than $10,000	○ $50,000-59,999
○ $10,000-14,999	○ $60,000-74,999
○ $15,000-19,999	○ $75,000-99,999
○ $20,000-24,999	○ $100,000-149,999
○ $25,000-29,999	○ $150,000-199,999
○ $30,000-39,999	○ $200,000-$249,999
○ $40,000-49,999	○ $250,000 or more

22. Current religious preference: (Mark one in each column)

	Yours	Father's	Mother's
Baptist	Ⓨ	Ⓕ	Ⓜ
Buddhist	Ⓨ	Ⓕ	Ⓜ
Eastern Orthodox	Ⓨ	Ⓕ	Ⓜ
Episcopal	Ⓨ	Ⓕ	Ⓜ
Islamic	Ⓨ	Ⓕ	Ⓜ
Jewish	Ⓨ	Ⓕ	Ⓜ
LDS (Mormon)	Ⓨ	Ⓕ	Ⓜ
Lutheran	Ⓨ	Ⓕ	Ⓜ
Methodist	Ⓨ	Ⓕ	Ⓜ
Presbyterian	Ⓨ	Ⓕ	Ⓜ
Quaker	Ⓨ	Ⓕ	Ⓜ
Roman Catholic	Ⓨ	Ⓕ	Ⓜ
Seventh Day Adventist	Ⓨ	Ⓕ	Ⓜ
United Church of Christ	Ⓨ	Ⓕ	Ⓜ
Other Christian	Ⓨ	Ⓕ	Ⓜ
Other Religion	Ⓨ	Ⓕ	Ⓜ
None	Ⓨ	Ⓕ	Ⓜ

23. My religion treats men and women as equals. (Mark one)

○ Always ○ Sometimes ○ Never ○ Not applicable

24. Please indicate the ethnic background of yourself, your father, and your mother. (Mark all that apply in each column)

	You	Father	Mother
White/Caucasian	Ⓨ	Ⓕ	Ⓜ
African American/Black	Ⓨ	Ⓕ	Ⓜ
American Indian/Alaska Native	Ⓨ	Ⓕ	Ⓜ
Asian American/Asian	Ⓨ	Ⓕ	Ⓜ
Native Hawaiian/Pacific Islander	Ⓨ	Ⓕ	Ⓜ
Mexican American/Chicano	Ⓨ	Ⓕ	Ⓜ
Puerto Rican	Ⓨ	Ⓕ	Ⓜ
Other Latino	Ⓨ	Ⓕ	Ⓜ
Other	Ⓨ	Ⓕ	Ⓜ

25. For the activities below, indicate which ones you did during the past year. If you engaged in an activity frequently, mark Ⓕ. If you engaged in an activity one or more times, but not frequently, mark Ⓞ (occasionally). Mark Ⓝ (Not at all) if you have not performed the activity during the past year. (Mark one for each item)

	Frequently	Occasionally	Not at all
Attended a religious service	Ⓕ	Ⓞ	Ⓝ
Was bored in class	Ⓕ	Ⓞ	Ⓝ
Participated in organized demonstrations	Ⓕ	Ⓞ	Ⓝ
Tutored another student	Ⓕ	Ⓞ	Ⓝ
Studied with other students	Ⓕ	Ⓞ	Ⓝ
Was a guest in a teacher's home	Ⓕ	Ⓞ	Ⓝ
Smoked cigarettes	Ⓕ	Ⓞ	Ⓝ
Drank beer	Ⓕ	Ⓞ	Ⓝ
Drank wine or liquor	Ⓕ	Ⓞ	Ⓝ
Felt overwhelmed by all I had to do	Ⓕ	Ⓞ	Ⓝ
Felt depressed	Ⓕ	Ⓞ	Ⓝ
Performed volunteer work	Ⓕ	Ⓞ	Ⓝ
Played a musical instrument	Ⓕ	Ⓞ	Ⓝ
Asked a teacher for advice after class	Ⓕ	Ⓞ	Ⓝ
Overslept and missed class or appointment	Ⓕ	Ⓞ	Ⓝ
Discussed politics	Ⓕ	Ⓞ	Ⓝ
Voted in a student election	Ⓕ	Ⓞ	Ⓝ
Socialized with someone of another racial/ethnic group	Ⓕ	Ⓞ	Ⓝ
Came late to class	Ⓕ	Ⓞ	Ⓝ
Attended a public recital or concert	Ⓕ	Ⓞ	Ⓝ
Visited an art gallery or museum	Ⓕ	Ⓞ	Ⓝ
Discussed religion	Ⓕ	Ⓞ	Ⓝ
Communicated via e-mail	Ⓕ	Ⓞ	Ⓝ
Used the Internet for research or homework	Ⓕ	Ⓞ	Ⓝ
Participated in Internet chat rooms	Ⓕ	Ⓞ	Ⓝ
Other Internet use	Ⓕ	Ⓞ	Ⓝ
Performed community service as part of a class	Ⓕ	Ⓞ	Ⓝ
Used a personal computer	Ⓕ	Ⓞ	Ⓝ

26. How many Advanced Placement courses or exams did you take in high school? (Mark one in each row)

	None	1	2-3	4-6	7-10	11+
AP Courses	○	○	○	○	○	○
AP Exams	○	○	○	○	○	○

27. What is the highest level of formal education obtained by your parents? (Mark one in each column)

	Father	Mother
Grammar school or less	○	○
Some high school	○	○
High school graduate	○	○
Postsecondary school other than college	○	○
Some college	○	○
College degree	○	○
Some graduate school	○	○
Graduate degree	○	○

28. How would you characterize your political views? (Mark one)

○ Far left
○ Liberal
○ Middle-of-the-road
○ Conservative
○ Far right

29. In deciding to go to college, how important to you was each of the following reasons? (Mark one answer for each possible reason)

	Very important	Somewhat important	Not important
My parents wanted me to go	Ⓥ	Ⓢ	Ⓝ
I could not find a job	Ⓥ	Ⓢ	Ⓝ
Wanted to get away from home	Ⓥ	Ⓢ	Ⓝ
To be able to get a better job	Ⓥ	Ⓢ	Ⓝ
To gain a general education and appreciation of ideas	Ⓥ	Ⓢ	Ⓝ
To improve my reading and study skills	Ⓥ	Ⓢ	Ⓝ
There was nothing better to do	Ⓥ	Ⓢ	Ⓝ
To make me a more cultured person	Ⓥ	Ⓢ	Ⓝ
To be able to make more money	Ⓥ	Ⓢ	Ⓝ
To learn more about things that interest me	Ⓥ	Ⓢ	Ⓝ
To prepare myself for graduate or professional school	Ⓥ	Ⓢ	Ⓝ
A mentor/role model encouraged me to go	Ⓥ	Ⓢ	Ⓝ
To get training for a specific career	Ⓥ	Ⓢ	Ⓝ

30. Rate yourself on each of the following traits as compared with the average person your age. We want the most accurate estimate of how you see yourself. (Mark one in each row)

	Highest 10%	Above Average	Average	Below Average	Lowest 10%
Academic ability	○	○	○	○	○
Artistic ability	○	○	○	○	○
Computer skills	○	○	○	○	○
Cooperativeness	○	○	○	○	○
Creativity	○	○	○	○	○
Drive to achieve	○	○	○	○	○
Emotional health	○	○	○	○	○
Leadership ability	○	○	○	○	○
Mathematical ability	○	○	○	○	○
Physical health	○	○	○	○	○
Persistence	○	○	○	○	○
Popularity	○	○	○	○	○
Public speaking ability	○	○	○	○	○
Religiousness	○	○	○	○	○
Risk-taking	○	○	○	○	○
Self-confidence (intellectual)	○	○	○	○	○
Self-confidence (social)	○	○	○	○	○
Self-understanding	○	○	○	○	○
Spirituality	○	○	○	○	○
Understanding of others	○	○	○	○	○
Writing ability	○	○	○	○	○

31. Mark only three responses, one in each column.

(M) Your mother's occupation
(F) Your father's occupation
(Y) Your probable career occupation

NOTE: If your father or mother is deceased, please indicate his or her last occupation.

	Y	F	M
Accountant or actuary	Y	F	M
Actor or entertainer	Y	F	M
Architect or urban planner	Y	F	M
Artist	Y	F	M
Business (clerical)	Y	F	M
Business executive (management, administrator)	Y	F	M
Business owner or proprietor	Y	F	M
Business salesperson or buyer	Y	F	M
Clergy (minister, priest)	Y	F	M
Clergy (other religious)	Y	F	M
Clinical psychologist	Y	F	M
College administrator/staff	Y	F	M
College teacher	Y	F	M
Computer programmer or analyst	Y	F	M
Conservationist or forester	Y	F	M
Dentist (including orthodontist)	Y	F	M
Dietitian or home economist	Y	F	M
Engineer	Y	F	M
Farmer or rancher	Y	F	M
Foreign service worker (including diplomat)	Y	F	M
Homemaker (full-time)	Y	F	M
Interior decorator (including designer)	Y	F	M
Lab technician or hygienist	Y	F	M
Law enforcement officer	Y	F	M
Lawyer (attorney) or judge	Y	F	M
Military service (career)	Y	F	M
Musician (performer, composer)	Y	F	M
Nurse	Y	F	M
Optometrist	Y	F	M
Pharmacist	Y	F	M
Physician	Y	F	M
Policymaker/Government	Y	F	M
School counselor	Y	F	M
School principal or superintendent	Y	F	M
Scientific researcher	Y	F	M
Social, welfare or recreation worker	Y	F	M
Therapist (physical, occupational speech)	Y	F	M
Teacher or administrator (elementary)	Y	F	M
Teacher or administrator (secondary)	Y	F	M
Veterinarian	Y	F	M
Writer or journalist	Y	F	M
Skilled trades	Y	F	M
Laborer (unskilled)	Y	F	M
Semi-skilled worker	Y	F	M
Unemployed	Y	F	M
Other	Y	F	M
Undecided	Y		

32. Mark one in each row:

(1) Disagree Strongly
(2) Disagree Somewhat
(3) Agree Somewhat
(4) Agree Strongly

	4	3	2	1
There is too much concern in the courts for the rights of criminals	4	3	2	1
Abortion should be legal	4	3	2	1
The death penalty should be abolished	4	3	2	1
Marijuana should be legalized	4	3	2	1
It is important to have laws prohibiting homosexual relationships	4	3	2	1
The federal government should do more to control the sale of handguns	4	3	2	1
Racial discrimination is no longer a major problem in America	4	3	2	1
Realistically, an individual can do little to bring about changes in our society	4	3	2	1
Wealthy people should pay a larger share of taxes than they do now	4	3	2	1
Colleges should prohibit racist/sexist speech on campus	4	3	2	1
Same-sex couples should have the right to legal marital status	4	3	2	1
Affirmative action in college admissions should be abolished	4	3	2	1
The activities of married women are best confined to the home and family	4	3	2	1
People should not obey laws which violate their personal values	4	3	2	1
Federal military spending should be increased	4	3	2	1
The federal government should do more to discourage energy consumption	4	3	2	1

33. Please indicate the extent to which each of the following describes you.
(Mark one for each item)

(1) Not at all
(2) To some extent
(3) To a great extent

	3	2	1
Searching for mission/purpose in life	3	2	1
Engaging in self-reflection	3	2	1
Appreciating the interconnectedness of everything	3	2	1
Believing in the sacredness of life	3	2	1
Being honest in my relationships with others	3	2	1

34. During your last year in high school, how much time did you spend during a typical week doing the following activities?

Hours per week:	None	Less than 1 hour	1-2	3-5	6-10	11-15	16-20	Over 20
Studying/homework								
Socializing with friends								
Talking with teachers outside of class								
Exercise or sports								
Partying								
Working (for pay)								
Volunteer work								
Student clubs/groups								
Watching TV								
Household/childcare duties								
Reading for pleasure								
Playing video/computer games								
Prayer/meditation								

35. Do you have any concern about your ability to finance your college education?
(Mark one)

- None (I am confident that I will have sufficient funds)
- Some (but I probably will have enough funds)
- Major (not sure I will have enough funds to complete college)

36. Below are some reasons that might have influenced your decision to attend this particular college. How important was each reason in your decision to come here?
(Mark one answer for each possible reason)

	Very Important	Somewhat Important	Not Important
My relatives wanted me to come here	V	S	N
My teacher advised me	V	S	N
This college has a very good academic reputation	V	S	N
This college has a good reputation for its social activities	V	S	N
I was offered financial assistance	V	S	N
This college offers special educational programs	V	S	N
This college has low tuition	V	S	N
High school counselor advised me	V	S	N
Private college counselor advised me	V	S	N
I wanted to live near home	V	S	N
Not offered aid by first choice	V	S	N
I was attracted by the religious affiliation/orientation of the college	V	S	N
I wanted to go to a school about the size of this college	V	S	N
Rankings in national magazines	V	S	N
Information from a website	V	S	N
I was admitted through an Early Action or Early Decision program	V	S	N
The athletic department recruited me	V	S	N
Reputation for campus safety	V	S	N

37. Below is a list of different undergraduate major fields grouped into general categories. Mark only one oval to indicate your probable field of study.

ARTS AND HUMANITIES
- Art, fine and applied ... 1
- English (language and literature) ... 2
- History ... 3
- Journalism ... 4
- Language and Literature (except English) ... 5
- Music ... 6
- Philosophy ... 7
- Speech ... 8
- Theater or Drama ... 9
- Theology or Religion ... 10
- Other Arts and Humanities ... 11

BIOLOGICAL SCIENCE
- Biology (general) ... 12
- Biochemistry or Biophysics ... 13
- Botany ... 14
- Environmental Science ... 15
- Marine (Life) Science ... 16
- Microbiology or Bacteriology ... 17
- Zoology ... 18
- Other Biological Science ... 19

BUSINESS
- Accounting ... 20
- Business Admin. (general) ... 21
- Finance ... 22
- International Business ... 23
- Marketing ... 24
- Management ... 25
- Secretarial Studies ... 26
- Other Business ... 27

EDUCATION
- Business Education ... 28
- Elementary Education ... 29
- Music or Art Education ... 30
- Physical Education or Recreation ... 31
- Secondary Education ... 32
- Special Education ... 33
- Other Education ... 34

ENGINEERING
- Aeronautical or Astronautical Eng. ... 35
- Civil Engineering ... 36
- Chemical Engineering ... 37
- Computer Engineering ... 38
- Electrical or Electronic Engineering ... 39
- Industrial Engineering ... 40
- Mechanical Engineering ... 41
- Other Engineering ... 42

PHYSICAL SCIENCE
- Astronomy ... 43
- Atmospheric Science (incl. Meteorology) ... 44
- Chemistry ... 45
- Earth Science ... 46
- Marine Science (incl. Oceanography) ... 47
- Mathematics ... 48
- Physics ... 49
- Statistics ... 50
- Other Physical Science ... 51

PROFESSIONAL
- Architecture or Urban Planning ... 52
- Home Economics ... 53
- Health Technology (medical, dental, laboratory) ... 54
- Library or Archival Science ... 55
- Medicine, Dentistry Veterinarian ... 56
- Nursing ... 57
- Pharmacy ... 58
- Therapy (occupational, physical, speech) ... 59
- Other Professional ... 60

SOCIAL SCIENCE
- Anthropology ... 61
- Economics ... 62
- Ethnic Studies ... 63
- Geography ... 64
- Political Science (gov't., international relations) ... 65
- Psychology ... 66
- Social Work ... 67
- Sociology ... 68
- Women's Studies ... 69
- Other Social Science ... 70

TECHNICAL
- Building Trades ... 71
- Data Processing or Computer Programming ... 72
- Drafting or Design ... 73
- Electronics ... 74
- Mechanics ... 75
- Other Technical ... 76

OTHER FIELDS
- Agriculture ... 77
- Communications ... 78
- Computer Science ... 79
- Forestry ... 80
- Kinesiology ... 81
- Law Enforcement ... 82
- Military Science ... 83
- Other Field ... 84
- Undecided ... 85

DO NOT WRITE IN THIS AREA

38. Please indicate the importance to you personally of each of the following: (Mark one for each item)
(N) Not Important (S) Somewhat Important (V) Very Important (E) Essential

- Becoming accomplished in one of the performing arts (acting, dancing, etc.) ... E V S N
- Becoming an authority in my field ... E V S N
- Obtaining recognition from my colleagues for contributions to my special field ... E V S N
- Influencing the political structure ... E V S N
- Influencing social values ... E V S N
- Raising a family ... E V S N
- Having administrative responsibility for the work of others ... E V S N
- Being very well off financially ... E V S N
- Helping others who are in difficulty ... E V S N
- Making a theoretical contribution to science ... E V S N
- Writing original works (poems, novels, short stories, etc.) ... E V S N
- Creating artistic work (painting, sculpture, decorating, etc.) ... E V S N
- Becoming successful in a business of my own ... E V S N
- Becoming involved in programs to clean up the environment ... E V S N
- Developing a meaningful philosophy of life ... E V S N
- Participating in a community action program ... E V S N
- Helping to promote racial understanding ... E V S N
- Keeping up to date with political affairs ... E V S N
- Becoming a community leader ... E V S N
- Integrating spirituality into my life ... E V S N
- Improving my understanding of other countries and cultures ... E V S N

39. What is your best guess as to the chances that you will: (Mark one for each item)
(N) No Chance (L) Very Little Chance (S) Some Chance (V) Very Good Chance

- Change major field? ... V S L N
- Change career choice? ... V S L N
- Participate in student government? ... V S L N
- Get a job to help pay for college expenses? ... V S L
- Work full-time while attending college? ... V S L N
- Join a social fraternity or sorority? ... V S L N
- Play varsity/intercollegiate athletics? ... V S L N
- Make at least a "B" average? ... V S L N
- Get a bachelor's degree (B.A., B.S., etc.)? ... V S L N
- Participate in student protests or demonstrations? ... V S L N
- Transfer to another college before graduating? ... V S L N
- Be satisfied with your college? ... V S L N
- Participate in volunteer or community service work? ... V S L N
- Seek personal counseling? ... V S L N
- Develop close friendships with other students? ... V S L N
- Communicate regularly with your professors? ... V S L N
- Socialize with someone of another racial/ethnic group? ... V S L N
- Participate in student clubs/groups? ... V S L N
- Strengthen religious beliefs/convictions? ... V S L N
- Participate in a study abroad program? ... V S L N
- Drop out of college? ... V S L N

40. Do you give the Higher Education Research Institute (HERI) permission to include your ID number should your college request the data for additional research analyses? HERI maintains strict standards of confidentiality and would require your college to sign a pledge of confidentiality. ○ Yes ○ No

The remaining ovals are provided for questions specifically designed by your college rather than the Higher Education Research Institute. If your college has chosen to use the ovals, please observe carefully the supplemental directions given to you.

41. A B C D E
42. A B C D E
43. A B C D E
44. A B C D E
45. A B C D E
46. A B C D E
47. A B C D E
48. A B C D E
49. A B C D E
50. A B C D E
51. A B C D E
52. A B C D E
53. A B C D E
54. A B C D E
55. A B C D E
56. A B C D E
57. A B C D E
58. A B C D E
59. A B C D E
60. A B C D E
61. A B C D E

THANK YOU!

R18103-Questar/20104C-03-54321

PLEASE PRINT (one letter or number per box)

NAME: FIRST | M I | LAST

When were you born? Month (01-12) Day (01-31) Year

ADDRESS:

STATE: ZIP: PHONE: – –

2002 COLLEGE STUDENT SURVEY

DIRECTIONS

Your responses will be read by an optical mark reader. Your careful observance of these few simple rules will be most appreciated.

- Use only black lead pencil (No. 2 is ideal).
- Make heavy black marks that fill the oval.
- Erase cleanly any answer you wish to change.
- Make no stray markings of any kind.

EXAMPLE:

Will marks made with ballpoint or felt-tip marker be properly read? Yes . . . ○ No . . . ●

FORM NO.: 213793

PLEASE PROVIDE YOUR SOCIAL SECURITY NO.

Mark here if directed: GRP. CODE A, GRP. CODE B

Dear Student:

This information is being collected as part of a continuing study of higher education conducted by the American Council on Education and the University of California at Los Angeles. Your participation in this research will help us to achieve a better understanding of how students are affected by their college experiences. Detailed information on this research program is available from the Higher Education Research Institute at UCLA. Identifying information has been requested in order to make subsequent mail follow-up studies possible. Your responses are held in the strictest professional confidence.

Sincerely,

Alexander W. Astin

Alexander W. Astin, Director
Higher Education Research Institute

PLEASE USE #2 PENCIL

1. What year did you first enter:
(Mark one in each column)

	Your 1st College	This College
2001 or 2002	○	○
2000	○	○
1999	○	○
1998	○	○
1997 or earlier	○	○

2. Your sex: ○ Male ○ Female

3. Please indicate the highest degree you (A) will have earned as of June 2002 and (B) plan to complete eventually at any institution.
(Mark one in each column)

	Highest Earned	Highest Planned
None	A	B
Vocational certificate	A	B
Associate (A.A. or equivalent)	A	B
Bachelor's degree (B.A., B.S., etc.)	A	B
Master's degree (M.A., M.S., etc.)	A	B
Ph.D. or Ed.D.	A	B
M.D., D.O., D.D.S., or D.V.M.	A	B
LL.B. or J.D. (Law)	A	B
B.D. or M.DIV. (Divinity)	A	B
Other	A	B

4. If you borrowed money to help pay for college expenses, estimate how much you will owe as of June 30, 2002:

$ [ ] [ ] [ ] [ ] [ ] [ ] .00

5. Please provide your test scores on the tests below: (If applicable)

GRE: Verbal . . . . . [ ] [ ] [ ]
GRE: Quantitative . . . . . [ ] [ ] [ ]
LSAT . . . . . [ ] [ ] [ ]
MCAT . . . . . [ ] [ ] [ ]
GMAT . . . . . [ ] [ ] [ ]

6. Since entering college have you:
(Mark all that apply)

- Joined a social fraternity or sorority. . . . ○
- Failed one or more courses. . . . . ○
- Had a part-time job on campus. . . . . ○
- Had a part-time job off campus. . . . . ○
- Worked full-time while attending school . ○
- Participated in student government . . . . ○
- Taken a remedial course . . . . . ○
- Taken an ethnic studies course . . . . . ○
- Taken a women's studies course . . . . . ○
- Attended a racial/cultural awareness workshop . . . . . ○
- Had a roommate of different race/ethnicity. ○
- Participated in an ethnic/racial student organization . . . . . ○
- Participated in:
  - intercollegiate football or basketball . . ○
  - other intercollegiate sport. . . . . ○
- Taken a leave of absence . . . . . ○
- Withdrawn from school . . . . . ○
- Been elected to student office . . . . . ○
- Enrolled in honors or advanced courses . ○
- Participated in an internship program . . ○
- Participated in leadership training . . . . . ○
- Transferred from a community college . . ○
- Transferred from a 4-year college . . . . . ○

7. Since entering college, indicate how often (Frequently, Occasionally or Not at all) you:
(Mark one for each item)

	Frequently	Occasionally	Not at all
Worked on independent study projects	F	O	N
Took interdisciplinary courses	F	O	N
Discussed course content with students outside of class	F	O	N
Worked on group projects in class	F	O	N
Have been a guest in a professor's home	F	O	N
Participated in intramural sports	F	O	N
Failed to complete homework on time	F	O	N
Felt bored in class	F	O	N
Studied with other students	F	O	N
Challenged a professor's ideas in class	F	O	N
Voted in a student election	F	O	N
Turned in course assignments electronically	F	O	N
Received course assignments through the Internet	F	O	N
Missed class due to employment	F	O	N
Tutored another college student	F	O	N
Felt supported by my family	F	O	N

8. If you could make your college choice over, would you still choose to enroll at your current (or most recent) college?

- ○ Definitely yes
- ○ Probably I would
- ○ Probably not
- ○ Definitely not
- ○ Don't know

DO NOT WRITE IN THIS BOX!

9. Please rate your satisfaction with your current (or most recent) college on each of the aspects of campus life listed below: (Mark one in each row)

	Very Satisfied	Satisfied	Neutral	Dissatisfied	Can't Rate/ No Experience
General education or core curriculum courses	V	S	N	D	?
Science and mathematics courses	V	S	N	D	?
Humanities courses	V	S	N	D	?
Social science courses	V	S	N	D	?
Courses in your major field	V	S	N	D	?
Relevance of coursework to everyday life	V	S	N	D	?
Overall quality of instruction	V	S	N	D	?
Laboratory facilities and equipment	V	S	N	D	?
Library facilities	V	S	N	D	?
Computer facilities	V	S	N	D	?
Quality of computer training/assistance	V	S	N	D	?
Availability of Internet access	V	S	N	D	?
Sense of community on campus	V	S	N	D	?
Tutoring or other academic assistance	V	S	N	D	?
Academic advising	V	S	N	D	?
Career counseling and advising	V	S	N	D	?
Student housing	V	S	N	D	?
Financial aid services	V	S	N	D	?
Amount of contact with faculty	V	S	N	D	?
Opportunities for community service	V	S	N	D	?
Job placement services for students	V	S	N	D	?
Campus health services	V	S	N	D	?
Class size	V	S	N	D	?
Interaction with other students	V	S	N	D	?
Ability to find a faculty or staff mentor	V	S	N	D	?
Leadership opportunities	V	S	N	D	?
Recreational facilities	V	S	N	D	?
Overall college experience	V	S	N	D	?

10. Please indicate your enrollment status below: (Mark one)

- ○ Full-time undergraduate
- ○ Part-time undergraduate
- ○ Graduate student
- ○ Not enrolled

11. Mark the one oval that best describes your undergraduate grade average.

- ○ A (3.75 - 4.0)
- ○ A-, B+ (3.25 - 3.74)
- ○ B (2.75 - 3.24)
- ○ B-, C+ (2.25 - 2.74)
- ○ C (1.75 - 2.24)
- ○ C- or less (below 1.75)

12. How would you characterize your political views? (Mark one)

- Far left ○
- Liberal ○
- Middle-of-the-road ○
- Conservative ○
- Far right ○

13. Please indicate the ethnic background of yourself, your father, and your mother. (Mark all that apply in each column)

	You	Father	Mother
White/Caucasian	○	○	○
African American/Black	○	○	○
American Indian/Alaska Native	○	○	○
Asian American/Asian	○	○	○
Native Hawaiian/Pacific Islander	○	○	○
Mexican American/Chicano	○	○	○
Puerto Rican	○	○	○
Other Latino	○	○	○
Other	○	○	○

14. Please mark your probable career/occupation below:

- Accountant or actuary ○
- Actor or entertainer ○
- Architect or urban planner ○
- Artist ○
- Business (clerical) ○
- Business executive (management, administrator) ○
- Business owner or proprietor ○
- Business salesperson or buyer ○
- Clergy (minister, priest) ○
- Clergy (other religious) ○
- Clinical psychologist ○
- College administrator/staff ○
- College teacher ○
- Computer programmer or analyst ○
- Conservationist or forester ○
- Dentist (including orthodontist) ○
- *Dietitian or home economist* ○
- Engineer ○
- Farmer or rancher ○
- Foreign service worker (including diplomat) ○
- Homemaker (full-time) ○
- *Interior decorator* (including designer) ○
- Lab technician or hygienist ○
- Law enforcement officer ○
- Lawyer (attorney) or judge ○
- Military service (career) ○
- Musician (performer, composer) ○
- Nurse ○
- Optometrist ○
- Pharmacist ○
- Physician ○
- Policymaker/government ○
- School counselor ○
- School principal or superintendent ○
- Scientific researcher ○
- Social, welfare or recreation worker ○
- Therapist (physical, occupational, speech) ○
- Teacher or administrator (elementary) ○
- Teacher or administrator (secondary) ○
- Veterinarian ○
- Writer or journalist ○
- Skilled trades ○
- Other ○
- Undecided ○

15. For the activities listed below, please indicate how often (Frequently, Occasionally, or Not at all) you engaged in each during the past year. (Mark one in each row)

	Frequently	Occasionally	Not at all
Smoked cigarettes	F	O	N
Felt lonely or homesick	F	O	N
Socialized with someone of another racial/ethnic group	F	O	N
Felt depressed	F	O	N
Felt overwhelmed by all I had to do	F	O	N
Attended a religious service	F	O	N
Drank beer	F	O	N
Drank wine or liquor	F	O	N
Performed volunteer work	F	O	N
Participated in organized demonstrations	F	O	N
Discussed politics	F	O	N
Overslept and missed class or appointment	F	O	N
Sought personal counseling	F	O	N
Visited an art gallery or museum	F	O	N
Discussed religion	F	O	N
Read the editorial page in the daily newspaper	F	O	N

16. During the past year, how much time did you spend during a typical week doing the following activities? (Mark one in each row)

	Hours Per Week							
	None	Less than 1 hour	1-2	3-5	6-10	11-15	16-20	Over 20
Studying/homework	○	○	○	○	○	○	○	○
Socializing with friends	○	○	○	○	○	○	○	○
Talking with faculty outside of class	○	○	○	○	○	○	○	○
Exercising/sports	○	○	○	○	○	○	○	○
Partying	○	○	○	○	○	○	○	○
Working (for pay)	○	○	○	○	○	○	○	○
Volunteer work	○	○	○	○	○	○	○	○
Student clubs/groups	○	○	○	○	○	○	○	○
Watching TV	○	○	○	○	○	○	○	○
Housework/childcare	○	○	○	○	○	○	○	○
Reading for pleasure	○	○	○	○	○	○	○	○
Using a personal computer	○	○	○	○	○	○	○	○
Commuting	○	○	○	○	○	○	○	○
Playing video games	○	○	○	○	○	○	○	○
Prayer/meditation	○	○	○	○	○	○	○	○
Classes/labs	○	○	○	○	○	○	○	○

17. What do you plan to be doing six months from now? (Mark all that apply)

- ○ Attending undergraduate college full-time
- ○ Attending undergraduate college part-time
- ○ Attending graduate/professional school
- ○ Working full-time
- ○ Working part-time
- ○ Participating in a community service organization
- ○ Serving in the Armed Forces
- ○ Attending a vocational training program
- ○ Traveling, hosteling, or backpacking
- ○ Doing volunteer work
- ○ Staying at home to be with or start a family
- ○ No current plans

18. Compared with when you first started college, how would you now describe your:

(Mark one for each item)

	Much Stronger	Stronger	No Change	Weaker	Much Weaker
General knowledge	5	4	3	2	1
Analytical and problem-solving skills	5	4	3	2	1
Knowledge of a particular field or discipline	5	4	3	2	1
Ability to think critically	5	4	3	2	1
Foreign language ability	5	4	3	2	1
Knowledge of people from different races/cultures	5	4	3	2	1
Religious beliefs and convictions	5	4	3	2	1
Leadership abilities	5	4	3	2	1
Interpersonal skills	5	4	3	2	1
Ability to get along with people of different races/cultures	5	4	3	2	1
Understanding of the problems facing your community	5	4	3	2	1
Understanding of social problems facing our nation	5	4	3	2	1
Writing skills	5	4	3	2	1
Public speaking ability	5	4	3	2	1
Ability to work cooperatively	5	4	3	2	1
Mathematical skills	5	4	3	2	1
Reading speed and comprehension	5	4	3	2	1
Computer skills	5	4	3	2	1

19. Indicate the importance to you personally of each of the following:

(Mark one for each item)

	Essential	Very Important	Somewhat Important	Not Important
Becoming accomplished in one of the performing arts (acting, dancing, etc.)	E	V	S	N
Becoming an authority in my field	E	V	S	N
Obtaining recognition from my colleagues for contributions to my special field	E	V	S	N
Influencing the political structure	E	V	S	N
Influencing social values	E	V	S	N
Raising a family	E	V	S	N
Having administrative responsibility for the work of others	E	V	S	N
Being very well off financially	E	V	S	N
Helping others who are in difficulty	E	V	S	N
Making a theoretical contribution to science	E	V	S	N
Writing original works (poems, novels, short stories, etc.)	E	V	S	N
Creating artistic work (painting, sculpture, decorating, etc.)	E	V	S	N
Becoming successful in a business of my own	E	V	S	N
Becoming involved in programs to clean up the environment	E	V	S	N
Developing a meaningful philosophy of life	E	V	S	N
Participating in a community action program	E	V	S	N
Helping to promote racial understanding	E	V	S	N
Keeping up to date with political affairs	E	V	S	N
Becoming a community leader	E	V	S	N

20. Your current religious preference: (Mark one)

Baptist	○	Presbyterian	○
Buddhist	○	Quaker	○
Eastern Orthodox	○	Roman Catholic	○
Episcopal	○	Seventh Day Adventist	○
Islamic	○	United Church of Christ	○
Jewish	○	Other Christian	○
LDS (Mormon)	○	Other Far Eastern Religion	○
Lutheran	○	Other Religion	○
Methodist	○	None	○

21. During the past year, how often did you:

(Mark one for each item)

	Daily	2 or 3 times/week	Once a week	1 or 2 times/month	Never
Communicate via e-mail:					
with faculty	5	4	3	2	1
with students at this college	5	4	3	2	1
with students at other colleges	5	4	3	2	1
with other friends or acquaintances	5	4	3	2	1
with your family	5	4	3	2	1
Participate in class discussions via e-mail/Internet	5	4	3	2	1
Use the Internet for research or homework	5	4	3	2	1
Use the Internet for nonacademic reasons	5	4	3	2	1

22. Do you own a personal computer? ○ Yes ○ No

23. Rate yourself on each of the following traits as compared with the average person your age. We want the most accurate estimate of how you see yourself.

(Mark one in each row)

	Highest 10%	Above Average	Average	Below Average	Lowest 10%
Academic ability	○	○	○	○	○
Artistic ability	○	○	○	○	○
Athletic ability	○	○	○	○	○
Competitiveness	○	○	○	○	○
Cooperativeness	○	○	○	○	○
Creativity	○	○	○	○	○
Drive to achieve	○	○	○	○	○
Emotional health	○	○	○	○	○
Leadership ability	○	○	○	○	○
Mathematical ability	○	○	○	○	○
Physical health	○	○	○	○	○
Popularity	○	○	○	○	○
Public speaking ability	○	○	○	○	○
Self-confidence (intellectual)	○	○	○	○	○
Self-confidence (social)	○	○	○	○	○
Self-understanding	○	○	○	○	○
Spirituality	○	○	○	○	○
Understanding of others	○	○	○	○	○
Writing ability	○	○	○	○	○
Religiousness/Religiosity	○	○	○	○	○

24. How often have professors at your current (or most recent) college provided you with:

(Mark one for each item)

	Frequently	Occasionally	Not at all
Encouragement to pursue graduate/professional study	○	○	○
An opportunity to work on a research project	○	○	○
Advice and guidance about your educational program	○	○	○
Respect (treated you like a colleague/peer)	○	○	○
An opportunity to publish	○	○	○
Emotional support and encouragement	○	○	○
A letter of recommendation	○	○	○
Assistance to improve your study skills	○	○	○
Negative feedback about your academic work	○	○	○
Intellectual challenge and stimulation	○	○	○
An opportunity to discuss coursework outside of class	○	○	○
Help in achieving your professional goals	○	○	○

25. Below is a list of different major fields.
(Mark only one in each column)

Ⓤ Undergraduate major (final or most recent)
Ⓖ Graduate major (omit if you do not plan to go to graduate school)

ARTS AND HUMANITIES
- Art, fine and applied ..... Ⓤ Ⓖ
- English (language and literature) ..... Ⓤ Ⓖ
- History ..... Ⓤ Ⓖ
- Journalism ..... Ⓤ Ⓖ
- Language and Literature (except English) ..... Ⓤ Ⓖ
- Music ..... Ⓤ Ⓖ
- Philosophy ..... Ⓤ Ⓖ
- Speech ..... Ⓤ Ⓖ
- Theater or Drama ..... Ⓤ Ⓖ
- Theology or Religion .... Ⓤ Ⓖ
- Other Arts and Humanities ..... Ⓤ Ⓖ

BIOLOGICAL SCIENCE
- Biology (general) ..... Ⓤ Ⓖ
- Biochemistry or Biophysics ..... Ⓤ Ⓖ
- Botany ..... Ⓤ Ⓖ
- Environmental Science... Ⓤ Ⓖ
- Marine (Life) Science .... Ⓤ Ⓖ
- Microbiology or Bacteriology ..... Ⓤ Ⓖ
- Zoology ..... Ⓤ Ⓖ
- Other Biological Science ..... Ⓤ Ⓖ

BUSINESS
- Accounting ..... Ⓤ Ⓖ
- Business Administration (general) ..... Ⓤ Ⓖ
- Finance ..... Ⓤ Ⓖ
- International Business ... Ⓤ Ⓖ
- Marketing ..... Ⓤ Ⓖ
- Management ..... Ⓤ Ⓖ
- Secretarial Studies ..... Ⓤ Ⓖ
- Other Business ..... Ⓤ Ⓖ

EDUCATION
- Business Education ..... Ⓤ Ⓖ
- Elementary Education ... Ⓤ Ⓖ
- Music or Art Education ... Ⓤ Ⓖ
- Physical Education or Recreation ..... Ⓤ Ⓖ
- Secondary Education .... Ⓤ Ⓖ
- Special Education ..... Ⓤ Ⓖ
- Other Education ..... Ⓤ Ⓖ

ENGINEERING
- Aero-/Astronautical Engineering ..... Ⓤ Ⓖ
- Civil Engineering..... Ⓤ Ⓖ
- Chemical Engineering ... Ⓤ Ⓖ
- Electrical or Electronic Engineering ..... Ⓤ Ⓖ
- Industrial Engineering.... Ⓤ Ⓖ
- Mechanical Engineering.. Ⓤ Ⓖ
- Other Engineering..... Ⓤ Ⓖ

PHYSICAL SCIENCE
- Astronomy ..... Ⓤ Ⓖ
- Atmospheric Science (incl. Meteorology) .... Ⓤ Ⓖ
- Chemistry ..... Ⓤ Ⓖ
- Earth Science ..... Ⓤ Ⓖ
- Marine Science (incl. Oceanography) ..... Ⓤ Ⓖ
- Mathematics ..... Ⓤ Ⓖ
- Physics ..... Ⓤ Ⓖ
- Statistics ..... Ⓤ Ⓖ
- Other Physical Science .. Ⓤ Ⓖ

PROFESSIONAL
- Architecture or Urban Planning ..... Ⓤ Ⓖ
- Home Economics ..... Ⓤ Ⓖ
- Health Technology (medical, dental, laboratory).. Ⓤ Ⓖ
- Law ..... Ⓖ
- Library/Archival Science . Ⓤ Ⓖ
- Medicine, Dentistry, Veterinarian ..... Ⓤ Ⓖ
- Nursing ..... Ⓤ Ⓖ
- Pharmacy ..... Ⓤ Ⓖ
- Therapy (occupational, physical, speech) ..... Ⓤ Ⓖ
- Other Professional ..... Ⓤ Ⓖ

SOCIAL SCIENCE
- Anthropology ..... Ⓤ Ⓖ
- Economics ..... Ⓤ Ⓖ
- Ethnic Studies ..... Ⓤ Ⓖ
- Geography..... Ⓤ Ⓖ
- Political Science (gov't., international relations).. Ⓤ Ⓖ
- Psychology ..... Ⓤ Ⓖ
- Social Work ..... Ⓤ Ⓖ
- Sociology ..... Ⓤ Ⓖ
- Women's Studies ..... Ⓤ Ⓖ
- Other Social Science .... Ⓤ Ⓖ

TECHNICAL
- Building Trades ..... Ⓤ Ⓖ
- Data Processing or Computer Programming Ⓤ Ⓖ
- Drafting or Design ..... Ⓤ Ⓖ
- Electronics ..... Ⓤ Ⓖ
- Mechanics ..... Ⓤ Ⓖ
- Other Technical ..... Ⓤ Ⓖ

OTHER FIELDS
- Agriculture ..... Ⓤ Ⓖ
- Communications (radio, TV, etc.) ..... Ⓤ Ⓖ
- Computer Science ..... Ⓤ Ⓖ
- Forestry ..... Ⓤ Ⓖ
- Law Enforcement ..... Ⓤ Ⓖ
- Military Science ..... Ⓤ Ⓖ
- Other Field ..... Ⓤ Ⓖ
- Undecided ..... Ⓤ Ⓖ

26. Please indicate your agreement with each of the following statements.
(Mark one for each item)

	Agree Strongly	Agree Somewhat	Disagree Somewhat	Disagree Strongly
Abortion should be legal	4	3	2	1
The death penalty should be abolished	4	3	2	1
If two people really like each other, it's all right for them to have sex even if they've known each other for only a very short time	4	3	2	1
The activities of married women are best confined to the home and family	4	3	2	1
Marijuana should be legalized	4	3	2	1
It is important to have laws prohibiting homosexual relationships	4	3	2	1
Racial discrimination is no longer a major problem in America	4	3	2	1
Realistically, an individual can do little to bring about changes in our society	4	3	2	1
Wealthy people should pay a larger share of taxes than they do now	4	3	2	1
Colleges should prohibit racist/sexist speech on campus	4	3	2	1
Affirmative action in college admissions should be abolished	4	3	2	1
There is too much concern in the courts for the rights of criminals	4	3	2	1
The federal government should do more to control the sale of handguns	4	3	2	1
Same sex couples should have the right to legal marital status	4	3	2	1
Federal military spending should be increased	4	3	2	1
All federal and state documents should be printed in English only	4	3	2	1
Material on the Internet should be regulated by the government	4	3	2	1

27. Is English your native language? ○ Yes ○ No

28. Since entering college, how many of your courses have included community service/service learning?
○ None (skip to question 31) ○ One ○ Two or more

29. In your most recent course that included service, how often did the professor:
(Mark one for each item)

	Frequently	Occasionally	Not at all
Encourage class discussions	F	O	N
Deliver lectures	F	O	N
Connect the service experience to the course material	F	O	N
Require written reflections of your service experience	F	O	N
How often did you:			
Apply the course material to your service work	F	O	N
Feel that the service experience increased your understanding of the academic course material	F	O	N
Feel that your service made a difference	F	O	N

30. In this most recent course, community service was:
○ Required ○ Optional

31. Do you give the Higher Education Research Institute at UCLA permission to include your ID number should your college request the data for additional research analyses? ○ Yes ○ No

ADDITIONAL QUESTIONS: If you received an additional page of questions, please mark your answers below:

32. Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	39. Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	46. Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ
33. Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	40. Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	47. Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ
34. Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	41. Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	48. Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ
35. Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	42. Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	49. Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ
36. Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	43. Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	50. Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ
37. Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	44. Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	51. Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ
38. Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	45. Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	**THANK YOU!**

2002-R17805-Questar/10744C-03-54321

PLEASE PRINT NAME AND PERMANENT/HOME ADDRESS (one letter or number per box)

NAME: FIRST MI LAST

When were you born? Month (01-12) Day (01-31) Year

ADDRESS:

CITY: STATE: ZIP: PHONE: – –

2006 CIRP FRESHMAN SURVEY

MARKING DIRECTIONS

Your responses will be read by an optical mark reader. Please,

- Use a pencil or black or blue pen.
- Fill in the oval completely.
- Erase cleanly any answer you wish to change or "X" out mark if in pen.

CORRECT MARK INCORRECT MARKS

PLEASE PROVIDE YOUR ID NUMBER (as instructed) | Mark here if directed | GROUP CODE A | GROUP CODE B

1. Your sex: Male Female

2. How old will you be on December 31 of this year? (Mark one)

16 or younger | 17 | 18 | 19 | 20 | 21-24 | 25-29 | 30-39 | 40-54 | 55 or older

3. Is English your native language? Yes No

4. In what year did you graduate from high school? (Mark one)

2006 | 2005 | 2004 | 2003 or earlier | Did not graduate but passed G.E.D. test | Never completed high school

5. Are you enrolled (or enrolling) as a: (Mark one) Full-time student? Part-time student?

6. How many miles is this college from your permanent home? (Mark one)

5 or less | 6-10 | 11-50 | 51-100 | 101-500 | Over 500

7. What was your average grade in high school? (Mark one)

A or A+ | A– | B+ | B | B– | C+ | C | D

8. What were your scores on the SAT I and/or ACT?

SAT VERBAL

SAT MATH

ACT Composite

9. Citizenship status:

U.S. citizen | Permanent resident (green card) | Neither

10. Prior to this term, have you ever taken courses for credit at this institution? Yes No

11. Since leaving high school, have you ever taken courses, whether for credit or not for credit, at any other institution (university, 4- or 2-year college, technical, vocational, or business school)? Yes No

12. Where do you plan to live during the fall term? (Mark one)

With my family or other relatives | Other private home, apartment, or room | College residence hall | Fraternity or sorority house | Other campus student housing | Other

13. Is this college your: (Mark one)

First choice? | Second choice? | Third choice? | Less than third choice?

14. If this college was not your first choice, were you accepted by your first choice college? Yes No

15. To how many colleges other than this one did you apply for admission this year?

None | 1 | 2 | 3 | 4 | 5 | 6 | 7-10 | 11 or more

16. Are your parents: (Mark one)

Both alive and living with each other? | Both alive, divorced or living apart? | One or both deceased?

17. During high school (grades 9-12) how many years did you study each of the following subjects? (Mark one for each item)

	None	1/2	1	2	3	4	5 or more
English							
Mathematics							
Foreign Language							
Physical Science							
Biological Science							
History/Am. Govt.							
Computer Science							
Arts and/or Music							

18. How many Advanced Placement courses or exams did you take in high school? (Mark one in each row)

	Not offered at my high school	None	1-4	5-9	10-14	15+
AP Courses						
AP Exams						

19. What is the highest academic degree that you intend to obtain? (Mark one in each column)

	Highest Planned	Highest Planned At This College
None		
Vocational certificate		
Associate (A.A. or equivalent)		
Bachelor's degree (B.A., B.S., etc.)		
Master's degree (M.A., M.S., etc.)		
Ph.D. or Ed.D.		
M.D., D.O., D.D.S., or D.V.M.		
J.D. (Law)		
B.D. or M.DIV. (Divinity)		
Other		

20. How would you describe the racial composition of the high school you last attended and the neighborhood where you grew up? (Mark one in each row)

	Completely non-White	Mostly non-White	Roughly half non-White	Mostly White	Completely White
High school I last attended					
Neighborhood where I grew up					

21. Did either of your parents or legal guardians attend the institution that you are now attending?

Neither | Mother or female legal guardian only | Father or male legal guardian only | Both

22. How much of your first year's educational expenses (room, board, tuition, and fees) do you expect to cover from each of the sources listed below? (Mark one answer for each possible source)

	None	Less than $1,000	$1,000 to 2,999	$3,000 to 5,999	$6,000 to 9,999	over $10,000
a. My Own or Family Resources						
Parents, other relatives or friends	○	○	○	○	○	○
Spouse	○	○	○	○	○	○
Savings from summer work	○	○	○	○	○	○
Other savings	○	○	○	○	○	○
Part-time job on campus	○	○	○	○	○	○
Part-time job off campus	○	○	○	○	○	○
Full-time job while in college	○	○	○	○	○	○
b. Aid Which Need Not Be Repaid						
Pell Grant	○	○	○	○	○	○
Supplemental Educational Opportunity Grant	○	○	○	○	○	○
State Scholarship or Grant						
Merit-based	○	○	○	○	○	○
Need-based	○	○	○	○	○	○
College Work-Study Grant	○	○	○	○	○	○
College Grant/Scholarship (other than above)	○	○	○	○	○	○
Other private grant	○	○	○	○	○	○
Government Aid						
GI military benefits	○	○	○	○	○	○
ROTC	○	○	○	○	○	○
Other Government Aid	○	○	○	○	○	○
c. Aid Which Must Be Repaid						
Stafford Loan (GSL)	○	○	○	○	○	○
Perkins Loan	○	○	○	○	○	○
Other College Loan	○	○	○	○	○	○
Other Loan	○	○	○	○	○	○
d. Other Than Above	○	○	○	○	○	○

23. How many individuals in your household are dependent on your parents for financial support? (Include yourself and your parents)

○ 1 ○ 4
○ 2 ○ 5
○ 3 ○ 6 or more

24. What is your best estimate of your parents' total income last year? Consider income from all sources before taxes. (Mark one)

○ Less than $10,000 ○ $50,000-59,999
○ $10,000-14,999 ○ $60,000-74,999
○ $15,000-19,999 ○ $75,000-99,999
○ $20,000-24,999 ○ $100,000-149,999
○ $25,000-29,999 ○ $150,000-199,999
○ $30,000-39,999 ○ $200,000-249,999
○ $40,000-49,999 ○ $250,000 or more

25. Do you have any concern about your ability to finance your college education? (Mark one)

○ None (I am confident that I will have sufficient funds)
○ Some (but I probably will have enough funds)
○ Major (not sure I will have enough funds to complete college)

26. Current religious preference: (Mark one in each column)

	Yours	Father's	Mother's
Baptist	Y	F	M
Buddhist	Y	F	M
Church of Christ	Y	F	M
Eastern Orthodox	Y	F	M
Episcopalian	Y	F	M
Hindu	Y	F	M
Islamic	Y	F	M
Jewish	Y	F	M
LDS (Mormon)	Y	F	M
Lutheran	Y	F	M
Methodist	Y	F	M
Presbyterian	Y	F	M
Quaker	Y	F	M
Roman Catholic	Y	F	M
Seventh Day Adventist	Y	F	M
United Church of Christ/ Congregational	Y	F	M
Other Christian	Y	F	M
Other Religion	Y	F	M
None	Y	F	M

27. For the activities below, indicate which ones you did during the past year. If you engaged in an activity frequently, mark (F). If you engaged in an activity one or more times, but not frequently, mark (O) (Occasionally). Mark (N) (Not at all) if you have not performed the activity during the past year. (Mark one for each item)

	Frequently	Occasionally	Not at all
Attended a religious service	F	O	N
Was bored in class	F	O	N
Participated in organized demonstrations	F	O	N
Tutored another student	F	O	N
Studied with other students	F	O	N
Was a guest in a teacher's home	F	O	N
Smoked cigarettes	F	O	N
Drank beer	F	O	N
Drank wine or liquor	F	O	N
Felt overwhelmed by all I had to do	F	O	N
Felt depressed	F	O	N
Performed volunteer work	F	O	N
Played a musical instrument	F	O	N
Asked a teacher for advice after class	F	O	N
Voted in a student election	F	O	N
Socialized with someone of another racial/ethnic group	F	O	N
Came late to class	F	O	N
Used the Internet:			
For research or homework	F	O	N
To read news sites	F	O	N
To read blogs	F	O	N
Performed community service as part of a class	F	O	N
Discussed religion	F	O	N
Discussed politics	F	O	N
Read a newspaper for:			
National and global news	F	O	N
Local news and information	F	O	N
Schoolwork	F	O	N

28. Rate yourself on each of the following traits as compared with the average person your age. We want the most accurate estimate of how you see yourself. (Mark one in each row)

	Highest 10%	Above Average	Average	Below Average	Lowest 10%
Academic ability	○	○	○	○	○
Artistic ability	○	○	○	○	○
Computer skills	○	○	○	○	○
Cooperativeness	○	○	○	○	○
Creativity	○	○	○	○	○
Drive to achieve	○	○	○	○	○
Emotional health	○	○	○	○	○
Leadership ability	○	○	○	○	○
Mathematical ability	○	○	○	○	○
Physical health	○	○	○	○	○
Public speaking ability	○	○	○	○	○
Religiousness	○	○	○	○	○
Self-confidence (intellectual)	○	○	○	○	○
Self-confidence (social)	○	○	○	○	○
Self-understanding	○	○	○	○	○
Spirituality	○	○	○	○	○
Understanding of others	○	○	○	○	○
Writing ability	○	○	○	○	○

29. What is the highest level of formal education obtained by your parents? (Mark one in each column)

	Father	Mother
Grammar school or less	○	○
Some high school	○	○
High school graduate	○	○
Postsecondary school other than college	○	○
Some college	○	○
College degree	○	○
Some graduate school	○	○
Graduate degree	○	○

30. In deciding to go to college, how important to you was each of the following reasons? (Mark one answer for each possible reason)

	Very Important	Somewhat Important	Not Important
My parents wanted me to go	V	S	N
I could not find a job	V	S	N
Wanted to get away from home	V	S	N
To be able to get a better job	V	S	N
To gain a general education and appreciation of ideas	V	S	N
There was nothing better to do	V	S	N
To make me a more cultured person	V	S	N
To be able to make more money	V	S	N
To learn more about things that interest me	V	S	N
To prepare myself for graduate or professional school	V	S	N
A mentor/role model encouraged me to go	V	S	N
To get training for a specific career	V	S	N

31. Mark only three responses, one in each column.

(M) Your mother's occupation
(F) Your father's occupation
(Y) Your probable career occupation

NOTE: If your father or mother is deceased, please indicate his or her last occupation.

	Y	F	M
Accountant or actuary	○	○	○
Actor or entertainer	○	○	○
Architect or urban planner	○	○	○
Artist	○	○	○
Business (clerical)	○	○	○
Business executive (management, administrator)	○	○	○
Business owner or proprietor	○	○	○
Business salesperson or buyer	○	○	○
Clergy (minister, priest)	○	○	○
Clergy (other religious)	○	○	○
Clinical psychologist	○	○	○
College administrator/staff	○	○	○
College teacher	○	○	○
Computer programmer or analyst	○	○	○
Conservationist or forester	○	○	○
Dentist (including orthodontist)	○	○	○
Dietitian or nutritionist	○	○	○
Engineer	○	○	○
Farmer or rancher	○	○	○
Foreign service worker (including diplomat)	○	○	○
Homemaker (full-time)	○	○	○
Interior decorator (including designer)	○	○	○
Lab technician or hygienist	○	○	○
Law enforcement officer	○	○	○
Lawyer (attorney) or judge	○	○	○
Military service (career)	○	○	○
Musician (performer, composer)	○	○	○
Nurse	○	○	○
Optometrist	○	○	○
Pharmacist	○	○	○
Physician	○	○	○
Policymaker/Government	○	○	○
School counselor	○	○	○
School principal or superintendent	○	○	○
Scientific researcher	○	○	○
Social, welfare, or recreation worker	○	○	○
Therapist (physical, occupational, speech)	○	○	○
Teacher or administrator (elementary)	○	○	○
Teacher or administrator (secondary)	○	○	○
Veterinarian	○	○	○
Writer or journalist	○	○	○
Skilled trades	○	○	○
Laborer (unskilled)	○	○	○
Semi-skilled worker	○	○	○
Unemployed	○	○	○
Other	○	○	○
Undecided	○		

32. Mark one in each row:

(1) Disagree Strongly
(2) Disagree Somewhat
(3) Agree Somewhat
(4) Agree Strongly

	4	3	2	1
There is too much concern in the courts for the rights of criminals	○	○	○	○
Abortion should be legal	○	○	○	○
The death penalty should be abolished	○	○	○	○
Marijuana should be legalized	○	○	○	○
It is important to have laws prohibiting homosexual relationships	○	○	○	○
Racial discrimination is no longer a major problem in America	○	○	○	○
Realistically, an individual can do little to bring about changes in our society	○	○	○	○
Wealthy people should pay a larger share of taxes than they do now	○	○	○	○
Same-sex couples should have the right to legal marital status	○	○	○	○
Affirmative action in college admissions should be abolished	○	○	○	○
Federal military spending should be increased	○	○	○	○
The federal government should do more to control the sale of handguns	○	○	○	○
Only volunteers should serve in the armed forces	○	○	○	○
The federal government is not doing enough to control environmental pollution	○	○	○	○
A national health care plan is needed to cover everybody's medical costs	○	○	○	○
Undocumented immigrants should be denied access to public education	○	○	○	○
Through hard work, everybody can succeed in American society	○	○	○	○
Dissent is a critical component of the political process	○	○	○	○
Colleges have the right to ban extreme speakers from campus	○	○	○	○
The chief benefit of a college education is that it increases one's earning power	○	○	○	○
The federal government should raise taxes to reduce the deficit	○	○	○	○

33. During your last year in high school, how much time did you spend during a typical week doing the following activities?

Hours per week:	None	Less than 1 hour	1-2	3-5	6-10	11-15	16-20	Over 20
Studying/homework	○	○	○	○	○	○	○	○
Socializing with friends	○	○	○	○	○	○	○	○
Talking with teachers outside of class	○	○	○	○	○	○	○	○
Exercise or sports	○	○	○	○	○	○	○	○
Partying	○	○	○	○	○	○	○	○
Working (for pay)	○	○	○	○	○	○	○	○
Volunteer work	○	○	○	○	○	○	○	○
Student clubs/groups	○	○	○	○	○	○	○	○
Watching TV	○	○	○	○	○	○	○	○
Household/childcare duties	○	○	○	○	○	○	○	○
Reading for pleasure	○	○	○	○	○	○	○	○
Playing video/ computer games	○	○	○	○	○	○	○	○

34. Are you: (Mark all that apply)

- ○ White/Caucasian
- ○ African American/Black
- ○ American Indian/Alaska Native
- ○ Asian American/Asian
- ○ Native Hawaiian/Pacific Islander
- ○ Mexican American/Chicano
- ○ Puerto Rican
- ○ Other Latino
- ○ Other

35. How would you characterize your political views? (Mark one)

- ○ Far left
- ○ Liberal
- ○ Middle-of-the-road
- ○ Conservative
- ○ Far right

36. Below are some reasons that might have influenced your decision to attend this particular college. How important was each reason in your decision to come here? (Mark one answer for each possible reason)

	Very Important	Somewhat Important	Not Important
My relatives wanted me to come here	(V)	(S)	(N)
My teacher advised me	(V)	(S)	(N)
This college has a very good academic reputation	(V)	(S)	(N)
This college has a good reputation for its social activities	(V)	(S)	(N)
I was offered financial assistance	(V)	(S)	(N)
The cost of attending this college	(V)	(S)	(N)
High school counselor advised me	(V)	(S)	(N)
Private college counselor advised me	(V)	(S)	(N)
I wanted to live near home	(V)	(S)	(N)
Not offered aid by first choice	(V)	(S)	(N)
Could not afford first choice	(V)	(S)	(N)
This college's graduates gain admission to top graduate/ professional schools	(V)	(S)	(N)
This college's graduates get good jobs	(V)	(S)	(N)
I was attracted by the religious affiliation/orientation of the college	(V)	(S)	(N)
I wanted to go to a school about the size of this college	(V)	(S)	(N)
Rankings in national magazines	(V)	(S)	(N)
Information from a website	(V)	(S)	(N)
I was admitted through an Early Action or Early Decision program	(V)	(S)	(N)
The athletic department recruited me	(V)	(S)	(N)
A visit to the campus	(V)	(S)	(N)

37. Below is a list of different undergraduate major fields grouped into general categories. Mark only one oval to indicate your probable field of study.

ARTS AND HUMANITIES
- Art, fine and applied (1)
- English (language and literature) (2)
- History (3)
- Journalism (4)
- Language and Literature (except English) (5)
- Music (6)
- Philosophy (7)
- Speech (8)
- Theater or Drama (9)
- Theology or Religion (10)
- Other Arts and Humanities (11)

BIOLOGICAL SCIENCE
- Biology (general) (12)
- Biochemistry or Biophysics (13)
- Botany (14)
- Environmental Science (15)
- Marine (Life) Science (16)
- Microbiology or Bacteriology (17)
- Zoology (18)
- Other Biological Science (19)

BUSINESS
- Accounting (20)
- Business Admin. (general) (21)
- Finance (22)
- International Business (23)
- Marketing (24)
- Management (25)
- Secretarial Studies (26)
- Other Business (27)

EDUCATION
- Business Education (28)
- Elementary Education (29)
- Music or Art Education (30)
- Physical Education or Recreation (31)
- Secondary Education (32)
- Special Education (33)
- Other Education (34)

ENGINEERING
- Aeronautical or Astronautical Eng (35)
- Civil Engineering (36)
- Chemical Engineering (37)
- Computer Engineering (38)
- Electrical or Electronic Engineering (39)
- Industrial Engineering (40)
- Mechanical Engineering (41)
- Other Engineering (42)

PHYSICAL SCIENCE
- Astronomy (43)
- Atmospheric Science (incl. Meteorology) (44)
- Chemistry (45)
- Earth Science (46)
- Marine Science (incl. Oceanography) (47)
- Mathematics (48)
- Physics (49)
- Statistics (50)
- Other Physical Science (51)

PROFESSIONAL
- Architecture or Urban Planning (52)
- Family & Consumer Sciences (53)
- Health Technology (medical, dental, laboratory) (54)
- Library or Archival Science (55)
- Medicine, Dentistry, Veterinary Medicine (56)
- Nursing (57)
- Pharmacy (58)
- Therapy (occupational, physical, speech) (59)
- Other Professional (60)

SOCIAL SCIENCE
- Anthropology (61)
- Economics (62)
- Ethnic Studies (63)
- Geography (64)
- Political Science (gov't., international relations) (65)
- Psychology (66)
- Social Work (67)
- Sociology (68)
- Women's Studies (69)
- Other Social Science (70)

TECHNICAL
- Building Trades (71)
- Data Processing or Computer Programming (72)
- Drafting or Design (73)
- Electronics (74)
- Mechanics (75)
- Other Technical (76)

OTHER FIELDS
- Agriculture (77)
- Communications (78)
- Computer Science (79)
- Forestry (80)
- Kinesiology (81)
- Law Enforcement (82)
- Military Science (83)
- Other Field (84)
- Undecided (85)

DO NOT WRITE IN THIS AREA

38. Please indicate the importance to you personally of each of the following: (Mark one for each item)

(N) Not Important (S) Somewhat Important (V) Very Important (E) Essential

	E	V	S	N
Becoming accomplished in one of the performing arts (acting, dancing, etc.)	E	V	S	N
Becoming an authority in my field	E	V	S	N
Obtaining recognition from my colleagues for contributions to my special field	E	V	S	N
Influencing the political structure	E	V	S	N
Influencing social values	E	V	S	N
Raising a family	E	V	S	N
Having administrative responsibility for the work of others	E	V	S	N
Being very well off financially	E	V	S	N
Helping others who are in difficulty	E	V	S	N
Making a theoretical contribution to science	E	V	S	N
Writing original works (poems, novels, short stories, etc.)	E	V	S	N
Creating artistic work (painting, sculpture, decorating, etc.)	E	V	S	N
Becoming successful in a business of my own	E	V	S	N
Becoming involved in programs to clean up the environment	E	V	S	N
Developing a meaningful philosophy of life	E	V	S	N
Participating in a community action program	E	V	S	N
Helping to promote racial understanding	E	V	S	N
Keeping up to date with political affairs	E	V	S	N
Becoming a community leader	E	V	S	N
Improving my understanding of other countries and cultures	E	V	S	N
Participating in an organization like the Peace Corps or AmeriCorps/VISTA	E	V	S	N

39. What is your best guess as to the chances that you will: (Mark one for each item)

(N) No Chance (L) Very Little Chance (S) Some Chance (V) Very Good Chance

	V	S	L	N
Change major field?	V	S	L	N
Change career choice?	V	S	L	N
Participate in student government?	V	S	L	N
Get a job to help pay for college expenses?	V	S	L	N
Work full-time while attending college?	V	S	L	N
Join a social fraternity or sorority?	V	S	L	N
Play varsity/intercollegiate athletics?	V	S	L	N
Make at least a "B" average?	V	S	L	N
Need extra time to complete your degree requirements?	V	S	L	N
Participate in student protests or demonstrations?	V	S	L	N
Transfer to another college before graduating?	V	S	L	N
Be satisfied with your college?	V	S	L	N
Participate in volunteer or community service work?	V	S	L	N
Seek personal counseling?	V	S	L	N
Communicate regularly with your professors?	V	S	L	N
Socialize with someone of another racial/ethnic group?	V	S	L	N
Participate in student clubs/groups?	V	S	L	N
Participate in a study abroad program?	V	S	L	N

40. Do you give the Higher Education Research Institute (HERI) permission to include your ID number should your college request the data for additional research analyses? HERI maintains strict standards of confidentiality and would require your college to sign a pledge of confidentiality. ○ Yes ○ No

The remaining ovals are provided for questions specifically designed by your college rather than the Higher Education Research Institute. If your college has chosen to use the ovals, please observe carefully the supplemental directions given to you.

41. A B C D E	47. A B C D E	53. A B C D E
42. A B C D E	48. A B C D E	54. A B C D E
43. A B C D E	49. A B C D E	55. A B C D E
44. A B C D E	50. A B C D E	56. A B C D E
45. A B C D E	51. A B C D E	57. A B C D E
46. A B C D E	52. A B C D E	58. A B C D E

THANK YOU!

Data Recognition Corp.-6G6044-6754-54321

PLEASE PRINT (one letter or number per box)

NAME: FIRST | M I | LAST

ADDRESS:

CITY: STATE: ZIP: PHONE: – –

When were you born? Month (01-12) Day (01-31) Year

2005-2006 COLLEGE STUDENT SURVEY

MARKING DIRECTIONS

Your responses will be read by an optical mark reader. Please,

- Use black or blue pen or a number 2 pencil.
- Make dark marks that fill the oval completely.
- Erase cleanly any mark you wish to change.
- Make no stray marks.

CORRECT MARK ⬬ INCORRECT MARKS ✓ ✗ ◒ ⊙

FORM NO:

PLEASE PROVIDE YOUR SOCIAL SECURITY NO.

Mark here if directed: GRP. CODE A, GRP. CODE B

1. What year did you first enter:

(Mark one in each column)

	Your 1st College	This College
2005 or 2006	⬭	⬭
2004	⬭	⬭
2003	⬭	⬭
2002	⬭	⬭
2001 or earlier	⬭	⬭

2. Your sex: ⬭ Male ⬭ Female

3. Is English your native language?

⬭ Yes ⬭ No

4. If you could make your college choice over, would you still choose to enroll at your current (or most recent) college?

⬭ Definitely yes
⬭ Probably I would
⬭ Probably not
⬭ Definitely no
⬭ Don't know

5. Please indicate the highest degree you (A) will have earned as of June 2006 and (B) plan to complete eventually at any institution.

(Mark one in each column)

	Highest Earned	Highest Planned
None	Ⓐ	Ⓑ
Vocational certificate	Ⓐ	Ⓑ
Associate (A.A. or equivalent)	Ⓐ	Ⓑ
Bachelor's degree (B.A., B.S., etc.)	Ⓐ	Ⓑ
Master's degree (M.A., M.S., etc.)	Ⓐ	Ⓑ
Ph.D or Ed. D.	Ⓐ	Ⓑ
M.D., D.O., D.D.S., or D.V.M.	Ⓐ	Ⓑ
LL.B. or J.D. (Law)	Ⓐ	Ⓑ
B.D. or M.DIV. (Divinity)	Ⓐ	Ⓑ
Other	Ⓐ	Ⓑ

6. If you borrowed money to help pay for college expenses, estimate how much you will owe as of June 30, 2006:

$ .00

7. Since entering college have you:

(Mark all that apply)

Joined a social fraternity or sorority ⬭
Failed one or more courses ⬭
Worked full-time while attending school ⬭
Participated in student government ⬭
Taken a remedial course ⬭
Taken an ethnic studies course ⬭
Taken a women's studies course ⬭
Attended a racial/cultural awareness workshop ⬭
Had a roommate of different race/ethnicity ⬭
Participated in an ethnic/racial student organization ⬭
Played varsity/intercollegiate athletics ⬭
Withdrew from school temporarily ⬭
Withdrew from school permanently ⬭
Enrolled in honors or advanced courses ⬭
Participated in an internship program ⬭
Participated in leadership training ⬭
Transferred from a community college ⬭
Transferred from a 4- year college ⬭
Participated in a study- abroad program ⬭

8. Since entering college, have you ever taken courses, whether for credit or not for credit, at any other institution (e.g., university, 4- or 2-year college, technical, vocational, or business school)?

⬭ Yes ⬭ No

9. Since entering college, indicate how often you:

(Mark one for each item)

	Frequently	Occasionally	Not at all
Worked on independent study projects	F	O	N
Discussed course content with students outside of class	F	O	N
Have been a guest in a professor's home	F	O	N
Participated in intramural sports	F	O	N
Failed to complete homework on time	F	O	N
Felt bored in class	F	O	N
Came late to class	F	O	N
Studied with other students	F	O	N
Performed community service as part of a class	F	O	N
Voted in a student election	F	O	N
Voted in a state/national election	F	O	N
Turned in course assignments electronically	F	O	N
Received course assignments through the Internet	F	O	N
Used the Internet for research or homework	F	O	N
Used the library for research or homework	F	O	N
Missed class due to employment	F	O	N
Tutored another college student	F	O	N
Met with an advisor/counselor about your career plans	F	O	N
Overslept and missed class or appointment	F	O	N
Fell asleep in class	F	O	N
Had difficulty getting the courses you needed	F	O	N
Used/purchased class notes from a professional service	F	O	N
Contested a grade	F	O	N
Were graded on a curve	F	O	N
Observed professors actively engaging students during class	F	O	N

PLEASE DO NOT WRITE IN THIS AREA

SERIAL #

10. Please rate your satisfaction with your current (or most recent) college in each area:
(Mark one in each row)

	Very Satisfied	Satisfied	Neutral	Dissatisfied	Can't Rate/ No Experience
General education or core curriculum courses	V	S	N	D	?
Science and mathematics courses	V	S	N	D	?
Humanities courses	V	S	N	D	?
Social science courses	V	S	N	D	?
Laboratory facilities and equipment	V	S	N	D	?
Library facilities	V	S	N	D	?
Computer facilities	V	S	N	D	?
Quality of computer training/assistance	V	S	N	D	?
Availability of Internet access	V	S	N	D	?
Tutoring or other academic assistance	V	S	N	D	?
Academic advising	V	S	N	D	?
Career counseling and advising	V	S	N	D	?
Student housing	V	S	N	D	?
Financial aid services	V	S	N	D	?
Opportunities for community service	V	S	N	D	?
Job placement services for students	V	S	N	D	?
Campus health services	V	S	N	D	?
Ability to find a faculty or staff mentor	V	S	N	D	?
Leadership opportunities	V	S	N	D	?
Recreational facilities	V	S	N	D	?

11. Please indicate your enrollment status below:
(Mark one)

- ○ Full- time undergraduate
- ○ Part- time undergraduate
- ○ Graduate student
- ○ Not enrolled

12. Mark the one oval that best describes your undergraduate grade average.

- ○ A (3.75 – 4.0)
- ○ A-, B + (3.25 – 3.74)
- ○ B (2.75 – 3.24)
- ○ B-, C+ (2.25 – 2.74)
- ○ C (1.75 – 2.24)
- ○ C- or less (below 1.75)

13. During the past year, how much time did you spend during a typical week doing the following activities?
(Mark one in each row)

	Hours Per Week							
	None	Less than 1 hour	1-2	3-5	6-10	11-15	16-20	Over 20
Studying/ homework	○	○	○	○	○	○	○	○
Attending classes/labs	○	○	○	○	○	○	○	○
Socializing with friends	○	○	○	○	○	○	○	○
Talking with faculty during office hours	○	○	○	○	○	○	○	○
Talking with faculty outside of class or office hours	○	○	○	○	○	○	○	○
Exercising/sports	○	○	○	○	○	○	○	○
Partying	○	○	○	○	○	○	○	○
Working (for pay) on campus	○	○	○	○	○	○	○	○
Working (for pay) off campus	○	○	○	○	○	○	○	○
Student clubs/ groups	○	○	○	○	○	○	○	○
Watching TV	○	○	○	○	○	○	○	○
Housework/childcare	○	○	○	○	○	○	○	○
Reading for pleasure	○	○	○	○	○	○	○	○
Commuting	○	○	○	○	○	○	○	○
Prayer/ meditation	○	○	○	○	○	○	○	○

14. Please mark your probable career/occupation below:
(Mark one)

- ○ Accountant or actuary
- ○ Actor or entertainer
- ○ Architect or urban planner
- ○ Artist
- ○ Business (clerical)
- ○ Business executive (management, administrator)
- ○ Business owner or proprietor
- ○ Business salesperson or buyer
- ○ Clergy (minister or priest)
- ○ Clergy (other religious)
- ○ Clinical psychologist
- ○ College administrator/staff
- ○ College teacher
- ○ Computer programmer or analyst
- ○ Conservationist or forester
- ○ Dentist (including orthodontist)
- ○ Dietitian or home economist
- ○ Engineer
- ○ Farmer or rancher
- ○ Foreign service worker (including diplomat)
- ○ Homemaker (full-time)
- ○ Interior decorator (including designer)
- ○ Lab technician or hygienist
- ○ Law enforcement officer
- ○ Lawyer (attorney) or judge
- ○ Military service (career)
- ○ Musician (performer, composer)
- ○ Nurse
- ○ Optometrist
- ○ Pharmacist
- ○ Physician
- ○ Policymaker/ government
- ○ School counselor
- ○ School principal or superintendent
- ○ Scientific researcher
- ○ Social, welfare or recreation worker
- ○ Therapist (physical, occupational, speech)
- ○ Teacher or administrator (elementary)
- ○ Teacher or administrator (secondary)
- ○ Veterinarian
- ○ Writer or journalist
- ○ Skilled trades
- ○ Other
- ○ Undecided

15. For the activities listed below, please indicate how often you engaged in each during the past year.
(Mark one in each row)

	Frequently	Occasionally	Not at all
Smoked cigarettes	F	O	N
Felt lonely or homesick	F	O	N
Socialized with someone of another racial/ethnic group	F	O	N
Felt depressed	F	O	N
Felt overwhelmed by all I had to do	F	O	N
Attended a religious service	F	O	N
Drank beer	F	O	N
Drank wine or liquor	F	O	N
Performed volunteer work	F	O	N
Participated in organized demonstrations	F	O	N
Discussed politics	F	O	N
Sought personal counseling	F	O	N
Discussed religion	F	O	N

16. How would you characterize your political views? (Mark one)

- ○ Far left
- ○ Liberal
- ○ Middle-of-the-road
- ○ Conservative
- ○ Far right

17. Please indicate your ethnic background.
(Mark all that apply)

- ○ White/ Caucasian
- ○ African American/ Black
- ○ American Indian/ Alaska Native
- ○ Asian American/ Asian
- ○ Native Hawaiian/ Pacific Islander
- ○ Mexican American/ Chicano
- ○ Puerto Rican
- ○ Other Latino
- ○ Other

18. Please rate your satisfaction with this institution on each of the aspects of campus life listed below.
(Mark one for each item)

	Very Satisfied	Satisfied	Neutral	Dissatisfied
Courses in your major field	V	S	N	D
Amount of contact with faculty	V	S	N	D
Class size	V	S	N	D
Interaction with other students	V	S	N	D
Relevance of coursework to everyday life	V	S	N	D
Relevance of coursework to future career plans	V	S	N	D
Overall quality of instruction	V	S	N	D
Overall sense of community among students	V	S	N	D
Availability of campus social activities	V	S	N	D
Overall college experience	V	S	N	D
Respect for the expression of diverse beliefs	V	S	N	D

19. Compared with when you first started college, how would you now describe your:

(Mark one for each item)

	Much Stronger	Stronger	No Change	Weaker	Much Weaker
General knowledge	5	4	3	2	1
Analytical and problem-solving skills	5	4	3	2	1
Knowledge of a particular field or discipline	5	4	3	2	1
Ability to think critically	5	4	3	2	1
Knowledge of people from different races/ cultures	5	4	3	2	1
Leadership abilities	5	4	3	2	1
Interpersonal skills	5	4	3	2	1
Ability to get along with people of different races/cultures	5	4	3	2	1
Understanding of the problems facing your community	5	4	3	2	1
Understanding of social problems facing our nation	5	4	3	2	1
Writing skills	5	4	3	2	1
Public speaking ability	5	4	3	2	1
Mathematical skills	5	4	3	2	1
Computer skills	5	4	3	2	1
Preparedness for employment after college	5	4	3	2	1
Preparedness for graduate or advanced education	5	4	3	2	1
Ability to manage your time effectively	5	4	3	2	1

20. Your current religious preference: (Mark one)

○ Baptist	○ Jewish	○ Roman Catholic
○ Buddhist	○ LDS (Mormon)	○ Seventh Day Adventist
○ Eastern Orthodox	○ Lutheran	○ United Church of Christ
○ Episcopal	○ Methodist	○ Other Christian
○ Hindu	○ Presbyterian	○ Other Religion
○ Islamic	○ Quaker	○ None

21. Indicate the importance to you personally of each of the following:

(Mark one for each item)

	Essential	Very Important	Somewhat Important	Not Important
Becoming accomplished in one of the performing arts (acting, dancing, etc.)	E	V	S	N
Becoming an authority in my field	E	V	S	N
Obtaining recognition from my colleagues for contributions to my special field	E	V	S	N
Influencing the political structure	E	V	S	N
Influencing the social values	E	V	S	N
Raising a family	E	V	S	N
Having administrative responsibility for the work of others	E	V	S	N
Being very well off financially	E	V	S	N
Helping others who are in difficulty	E	V	S	N
Making a theoretical contribution to science	E	V	S	N
Writing original works (poems, novels, etc.)	E	V	S	N
Creating artistic work (painting, sculpture, etc.)	E	V	S	N
Becoming successful in a business of my own	E	V	S	N
Becoming involved in programs to clean up the environment	E	V	S	N
Developing a meaningful philosophy of life	E	V	S	N
Participating in a community action program	E	V	S	N
Helping to promote racial understanding	E	V	S	N
Keeping up to date with political affairs	E	V	S	N
Becoming a community leader	E	V	S	N

22. Rate the climate of your college in general on the following continuum by filling in the appropriate oval.

(Mark one for each item set)

Friendly	5 4 3 2 1	Hostile
Socially Inclusive	5 4 3 2 1	Socially Exclusive
Intolerant of Diversity	5 4 3 2 1	Accepting of Diversity
Nonsexist	5 4 3 2 1	Sexist
Individualistic	5 4 3 2 1	Conformist
Impersonal	5 4 3 2 1	Caring
Improving	5 4 3 2 1	Worsening
Conservative	5 4 3 2 1	Liberal
Non-homophobic	5 4 3 2 1	Homophobic

23. Rate yourself on each of the following traits as compared with the average person your age. We want the most accurate estimate of how you see yourself.

(Mark one in each row)

	Highest 10%	Above Average	Average	Below Average	Lowest 10%
Academic ability	○	○	○	○	○
Artistic ability	○	○	○	○	○
Computer skills	○	○	○	○	○
Cooperativeness	○	○	○	○	○
Creativity	○	○	○	○	○
Drive to achieve	○	○	○	○	○
Emotional health	○	○	○	○	○
Leadership ability	○	○	○	○	○
Mathematical ability	○	○	○	○	○
Physical health	○	○	○	○	○
Persistence	○	○	○	○	○
Public speaking ability	○	○	○	○	○
Risk-taking	○	○	○	○	○
Self-confidence (intellectual)	○	○	○	○	○
Self-confidence (social)	○	○	○	○	○
Self-understanding	○	○	○	○	○
Spirituality	○	○	○	○	○
Understanding of others	○	○	○	○	○
Writing ability	○	○	○	○	○

24. How often have professors at your current (or more recent) college provided you with:

(Mark one for each item)

	Frequently	Occasionally	Not at all
Encouragement to pursue graduate/professional study	F	O	N
An opportunity to work on a research project	F	O	N
Advice and guidance about your educational program	F	O	N
Emotional support and encouragement	F	O	N
A letter of recommendation	F	O	N
Help to improve your study skills	F	O	N
Feedback on your academic work (outside of grades)	F	O	N
Intellectual challenge and stimulation	F	O	N
An opportunity to discuss coursework outside of class	F	O	N
Help in achieving your professional goals	F	O	N
An opportunity to apply classroom learning to "real-life" issues	F	O	N

EliteView™ forms by Pearson NCS MM262656-1 321 Printed in U.S.A.

25. Below is a list of different major fields. (Mark only one in each column)

Ⓤ Undergraduate major (final or most recent)

Ⓖ Graduate major (omit if you do not plan to go to graduate school)

ARTS AND HUMANITIES
- Art, fine and applied . . Ⓤ Ⓖ
- English (language and literature) . . . . . Ⓤ Ⓖ
- History . . . . . Ⓤ Ⓖ
- Journalism . . . . . Ⓤ Ⓖ
- Language and Literature (except English) . . . . Ⓤ Ⓖ
- Music . . . . . Ⓤ Ⓖ
- Philosophy . . . . . Ⓤ Ⓖ
- Speech . . . . . Ⓤ Ⓖ
- Theater or Drama . . . . Ⓤ Ⓖ
- Theology or Religion . . Ⓤ Ⓖ
- Other Arts and Humanities Ⓤ Ⓖ

BIOLOGICAL SCIENCE
- Biology (general) . . . . . Ⓤ Ⓖ
- Biochemistry or Biophysics Ⓤ Ⓖ
- Botany . . . . . Ⓤ Ⓖ
- Environmental Science Ⓤ Ⓖ
- Marine (Life) Science . . Ⓤ Ⓖ
- Microbiology or Bacteriology Ⓤ Ⓖ
- Zoology . . . . . Ⓤ Ⓖ
- Other Biological Science Ⓤ Ⓖ

BUSINESS
- Accounting . . . . . Ⓤ Ⓖ
- Business Admin. (general) Ⓤ Ⓖ
- Finance . . . . . Ⓤ Ⓖ
- International Business Ⓤ Ⓖ
- Marketing . . . . . Ⓤ Ⓖ
- Management . . . . . Ⓤ Ⓖ
- Secretarial Studies . . . . Ⓤ Ⓖ
- Other Business . . . . . Ⓤ Ⓖ

EDUCATION
- Business Education . . Ⓤ Ⓖ
- Elementary Education Ⓤ Ⓖ
- Music or Art Education Ⓤ Ⓖ
- Physical Education or Recreation . . . . . Ⓤ Ⓖ
- Secondary Education . . Ⓤ Ⓖ
- Special Education . . . . Ⓤ Ⓖ
- Other Education . . . . . Ⓤ Ⓖ

ENGINEERING
- Aeronautical or astronautical eng . . . . Ⓤ Ⓖ
- Civil Engineering . . . . . Ⓤ Ⓖ
- Chemical Engineering Ⓤ Ⓖ
- Computer Engineering Ⓤ Ⓖ
- Electrical or Electronic Engineering . . . . . Ⓤ Ⓖ
- Industrial Engineering . . Ⓤ Ⓖ
- Mechanical Engineering Ⓤ Ⓖ
- Other Engineering . . . . Ⓤ Ⓖ

PHYSICAL SCIENCE
- Astronomy . . . . . Ⓤ Ⓖ
- Atmospheric Science (incl. Meteorology) . . Ⓤ Ⓖ
- Chemistry . . . . . Ⓤ Ⓖ
- Earth Science . . . . . Ⓤ Ⓖ
- Marine Science (incl. Oceanography) Ⓤ Ⓖ
- Mathematics . . . . . Ⓤ Ⓖ
- Physics . . . . . Ⓤ Ⓖ
- Statistics . . . . . Ⓤ Ⓖ
- Other Physical Science Ⓤ Ⓖ

PROFESSIONAL
- Architecture or Urban Planning . . . . . Ⓤ Ⓖ
- Home Economics . . . . Ⓤ Ⓖ
- Health Technology (medical, dental, laboratory) . . . . . Ⓤ Ⓖ
- Law . . . . . Ⓖ
- Library/Archival Science Ⓤ Ⓖ
- Medicine, Dentistry, Veterinarian . . . . . Ⓤ Ⓖ
- Nursing . . . . . Ⓤ Ⓖ
- Pharmacy . . . . . Ⓤ Ⓖ
- Therapy (occupational, physical, speech) . . . . Ⓤ Ⓖ
- Other Professional . . . . Ⓤ Ⓖ

SOCIAL SCIENCE
- Anthropology . . . . . Ⓤ Ⓖ
- Economics . . . . . Ⓤ Ⓖ
- Ethnic Studies . . . . . Ⓤ Ⓖ
- Geography . . . . . Ⓤ Ⓖ
- Political Science (gov't. international relations) Ⓤ Ⓖ
- Psychology . . . . . Ⓤ Ⓖ
- Social Work . . . . . Ⓤ Ⓖ
- Sociology . . . . . Ⓤ Ⓖ
- Women's Studies . . . . Ⓤ Ⓖ
- Other Social Science . . Ⓤ Ⓖ

TECHNICAL
- Building Trades . . . . . Ⓤ Ⓖ
- Data Processing or Computer Programming Ⓤ Ⓖ
- Drafting or Design . . . . Ⓤ Ⓖ
- Electronics . . . . . Ⓤ Ⓖ
- Mechanics . . . . . Ⓤ Ⓖ
- Other Technical . . . . . Ⓤ Ⓖ

OTHER FIELDS
- Agriculture . . . . . Ⓤ Ⓖ
- Communications . . . . . Ⓤ Ⓖ
- Computer Science . . . . Ⓤ Ⓖ
- Forestry . . . . . Ⓤ Ⓖ
- Kinesiology . . . . . Ⓤ Ⓖ
- Law Enforcement . . . . Ⓤ Ⓖ
- Military Science . . . . . Ⓤ Ⓖ
- Other Field . . . . . Ⓤ Ⓖ
- Undecided . . . . . Ⓤ Ⓖ

26. Please indicate your agreement with each of the following statements.

(Mark one for each item)

	Agree Strongly	Agree Somewhat	Disagree Somewhat	Disagree Strongly
There is too much concern in the courts for the rights of criminals	4	3	2	1
Abortion should be legal	4	3	2	1
The death penalty should be abolished	4	3	2	1
Marijuana should be legalized	4	3	2	1
It is important to have laws prohibiting homosexual relationships	4	3	2	1
The federal government should do more to control the sale of handguns	4	3	2	1
Racial discrimination is no longer a major problem in America	4	3	2	1
Realistically, an individual can do little to bring about changes in our society	4	3	2	1
Wealthy people should pay a larger share of taxes than they do now	4	3	2	1
Colleges should prohibit racist/ sexist speech on campus	4	3	2	1
Same-sex couples should have the right to legal marital status	4	3	2	1
Affirmative action in college admissions should be abolished	4	3	2	1
The activities of married women are best confined to the home and family	4	3	2	1
Federal military spending should be increased	4	3	2	1
The federal government should do more to discourage energy consumption	4	3	2	1
The chief benefit of a college education is that it increases one's earning power	4	3	2	1

27. What do you plan to be doing in fall 2006? (Mark all that apply)
- ⬭ Attending undergraduate college full-time
- ⬭ Attending undergraduate college part-time
- ⬭ Attending graduate/ professional school
- ⬭ Working full-time
- ⬭ Working part-time
- ⬭ Participating in a community service organization
- ⬭ Serving in the Armed Forces
- ⬭ Attending a vocational training program
- ⬭ Traveling
- ⬭ Doing volunteer work
- ⬭ Staying at home to be with or start a family
- ⬭ No current plans

28. Do you give the Higher Education Research Institute (HERI) permission to include your ID number should your college request the data for additional research analyses? HERI maintains strict standards of confidentiality and would require your college to sign a pledge of confidentiality. ⬭ Yes ⬭ No

ADDITIONAL QUESTIONS: If you received an additional page of questions, please mark your answers below:

29.	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	39.	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	49.	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ
30.	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	40.	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	50.	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ
31.	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	41.	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	51.	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ
32.	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	42.	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	52.	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ
33.	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	43.	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	53.	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ
34.	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	44.	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	54.	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ
35.	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	45.	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	55.	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ
36.	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	46.	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	56.	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ
37.	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	47.	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	57.	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ
38.	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	48.	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	58.	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ

THANK YOU!

⬭⬭⬭⬭⬭⬭⬭⬭⬭⬭⬭⬭⬭⬭⬭⬭⬭⬭⬭⬭⬭⬭⬭⬭⬭ **SERIAL #**

PLEASE DO NOT WRITE IN THIS AREA

APPENDIX B. SELECTED RESULTS FROM THE RASCH PARTIAL CREDIT MODEL

Figure B.1

Rasch Item Map on the Civic Engagement Scale (without Expected Score Zones)

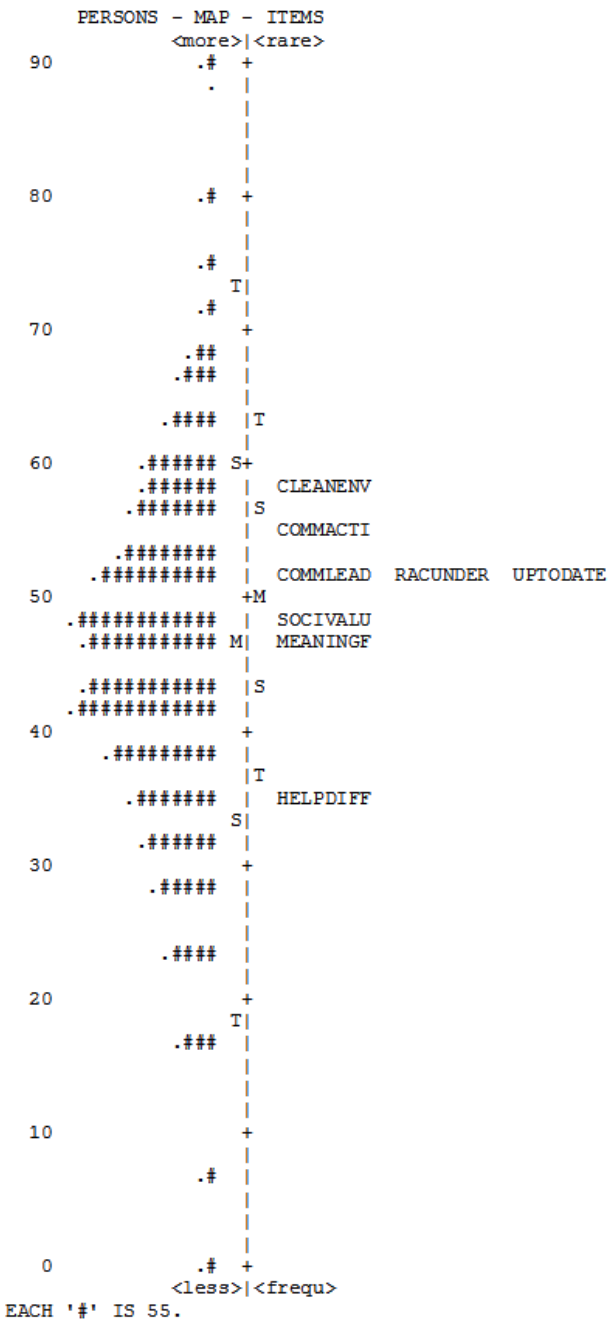


Figure B.2

Rasch Item Map on the Civic Engagement Scale (with Expected Score Zones)

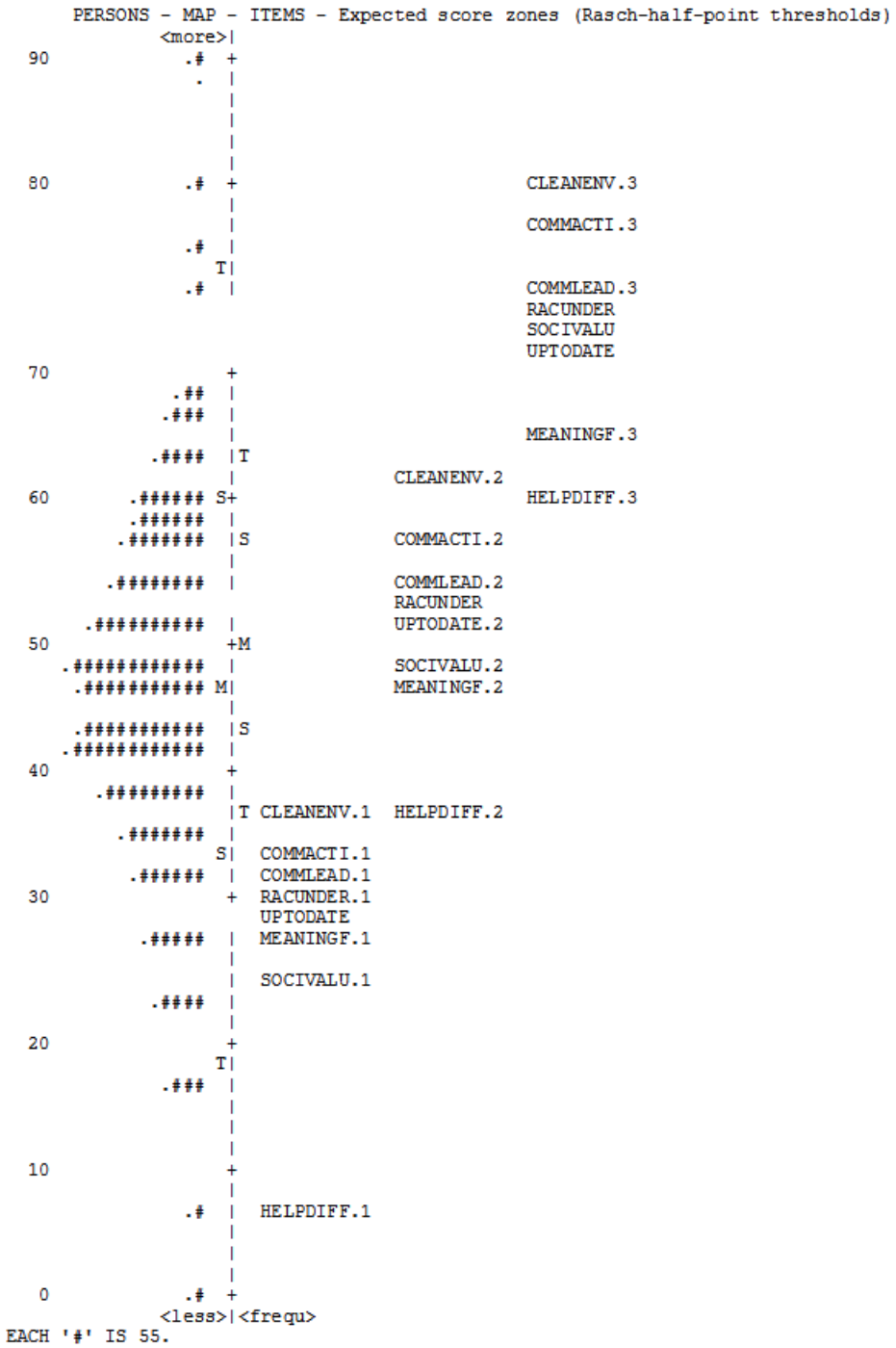


Figure B.3

Category Probability Curves for the Eight Items on the Civic Engagement Scale

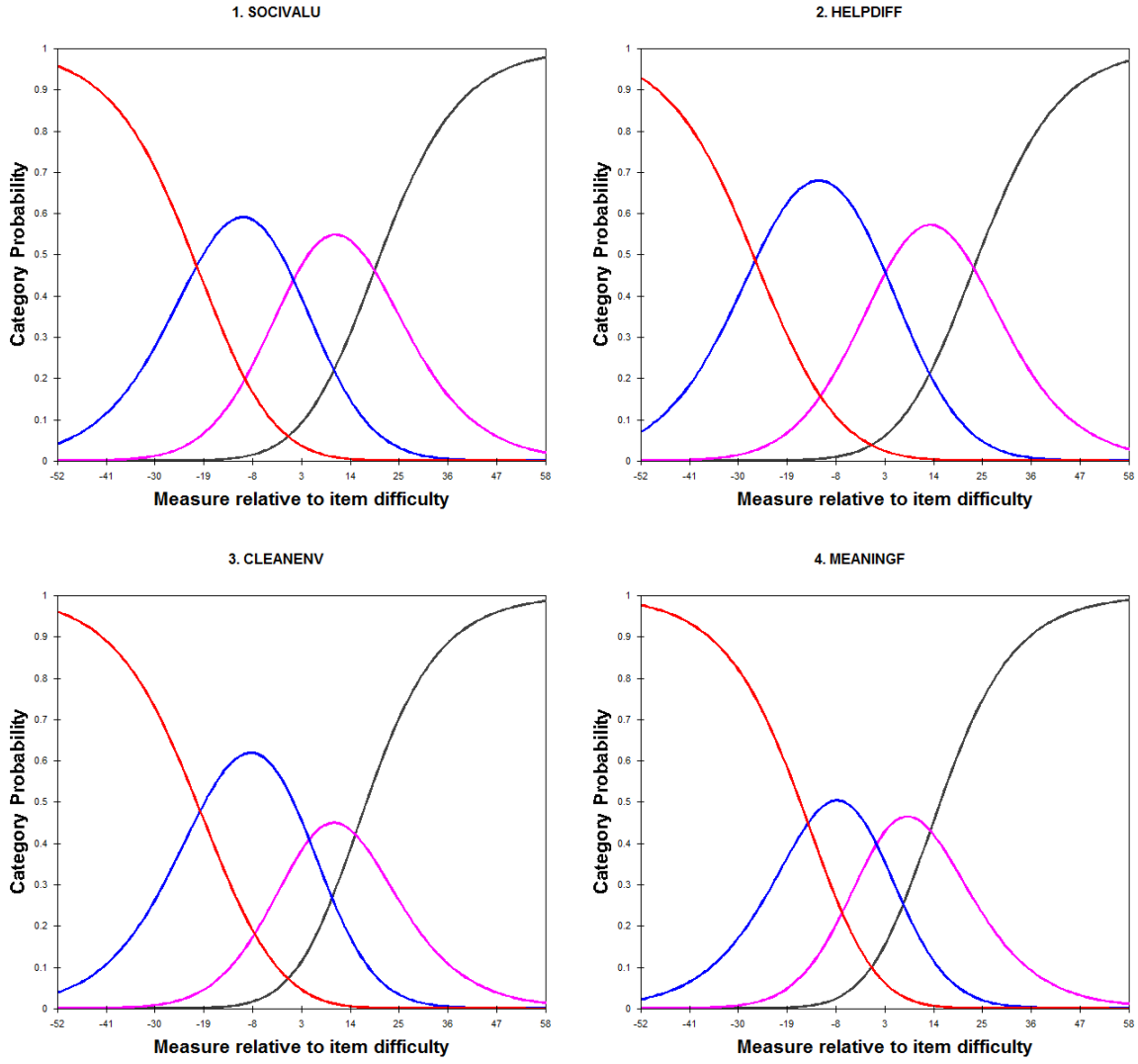
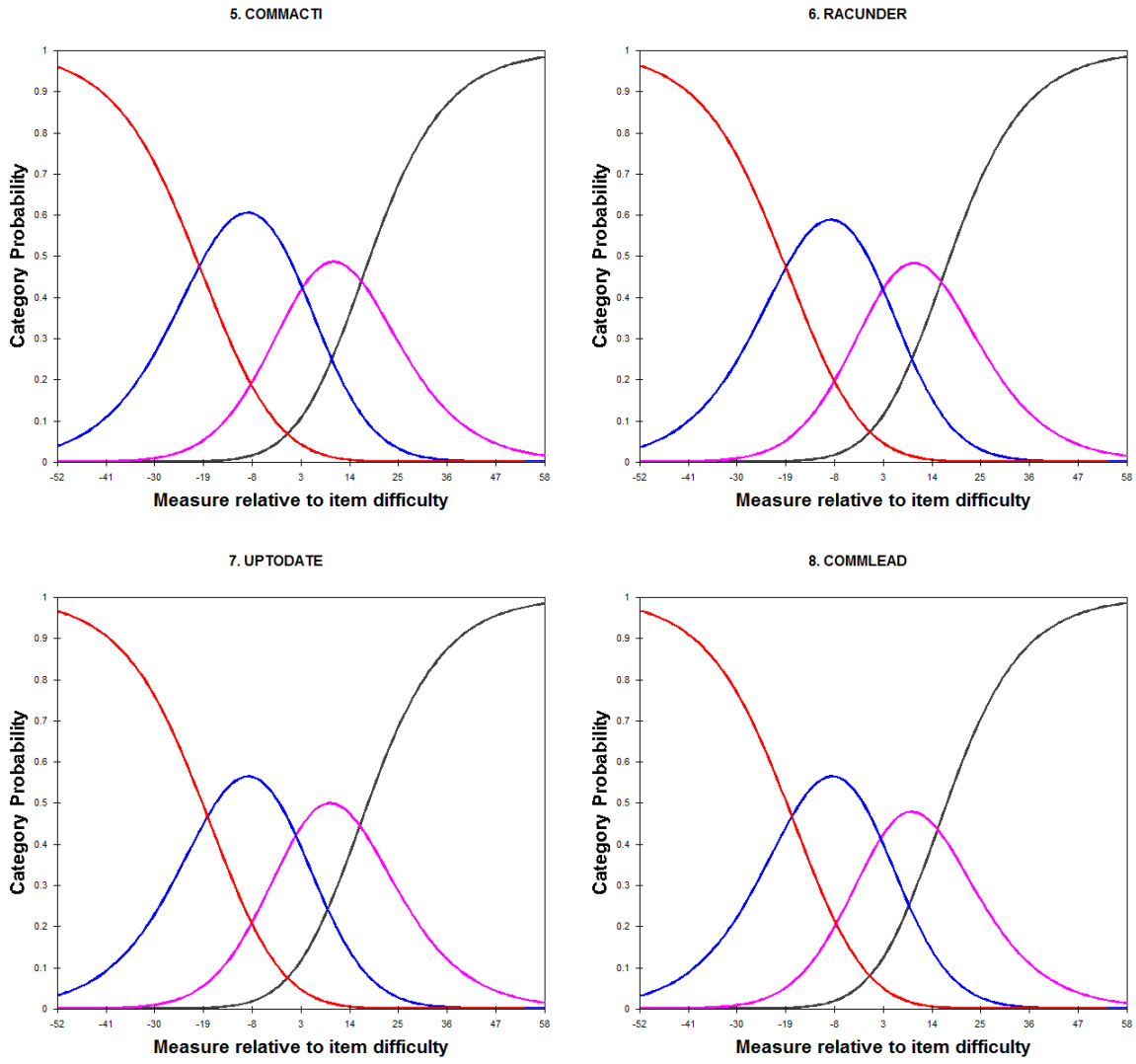


Figure B.3 (cont.)



APPENDIX C. ASSUMPTIONS IN THE OLS AND HLM ANALYSES IN RESEARCH QUESTION THREE

This appendix lists the results from testing the assumptions in the ordinary least squares (OLS) linear regression models and hierarchical linear models (HLMs) in the three value-added methods used in answering research question three.

Figures C.1-C.4 are results from testing the OLS assumptions in Method 1. Specifically, graphs (a) and (b) test the normality of the residuals of the regression model; graphs (c) and (d) test the linearity of the relationship between dependent and independent variables; graph (d) also tests the homoscedasticity of the error variance in the regression models. In all four regression models, results show that the normality, linearity, and homoscedasticity assumptions are not seriously violated. In graph (d), the residuals are not a function of the predicted values. However, they are not clustered closely around the horizontal line either. This means that some institutions are potential outliers in these OLS regression models. For the purpose of comparing different value-added methods, all institutions were included in the OLS analyses.

Tables C.1-C.2 and Figures C.5-C.6 are results from testing the HLM assumptions in Method 2. Tables C.1 (for the 2002 analysis) and C.2 (for the 2006 analysis) are results from the homogeneity of level-1 variance test. The null hypothesis of homogeneous level-1 variances across institutions are rejected in both analyses ($p < .001$). This heterogeneity may result from the omission of some important level-1 predictor variables, as some significant student covariates were found in research

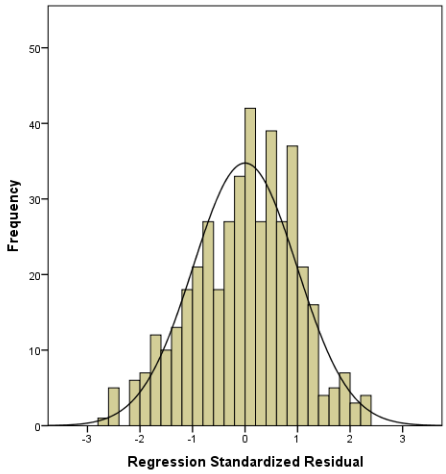
question one but they were not included in research question 3 because those items were not available in both freshman and senior surveys (e.g., SAT/ACT scores were not asked in the senior survey). As Raudenbush and Bryk (2002) pointed out, "a violation of the homogeneity assumption is not per se a serious problem for estimating either the level-2 coefficients or their standard errors," but the heterogeneity statistics are "quite sensitive to violation of normality in the observed data" (p.264-265). Therefore, the normality assumption was examined and results are shown in Figures C.5 and C.6.

Figures C.5 and C.6 are scatterplots of two summary statistics provided in the HLM level-2 residual file. The Mahalanobis distance provides a summary of the empirical Bayes residuals, u^*_{pj}, or the distance between the empirical Bayes estimates of the fixed effect parameters, β^*_{pj}, and the predicted values based on the level-2 equations, $\hat{\gamma}_{p0} + \sum \hat{\gamma}_{p0} W_{qj}$. Under the normality assumption, the Mahalanobis distances should be distributed approximately $\chi^2_{(v)}$, where *v* is the number of random effects. Similar to a univariate normal probability plot, normality of the random effects is satisfied if the plot of Mahalanobis distance and the expected values for a sample of *J* institutions resembles a 45 degree line (Raudenbush, Bryk, Cheong, Congdon, & Toit, 2004). In addition, Figures C.5 and C.6 are also helpful in detecting outliers. Again, some potential outliers are detected, but all institutions remained in the HLM analyses for the purpose of comparison.

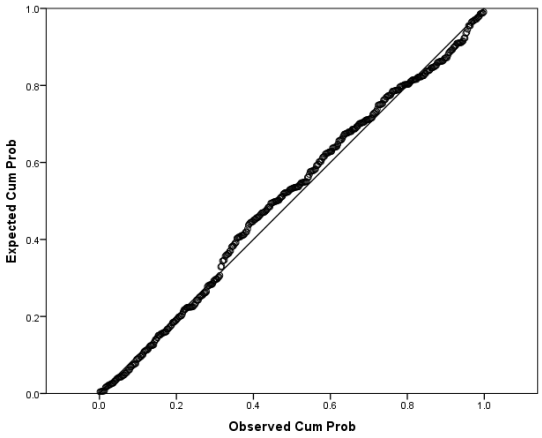
Similarly, Table C.3 and Figure C.7 are results from testing the HLM assumptions in Method 3. The null hypothesis of homogeneous level-1 variances across institutions

are rejected in Table C.3, but Figure C.7 shows less deviation from the normality, in contrast with earlier figures.

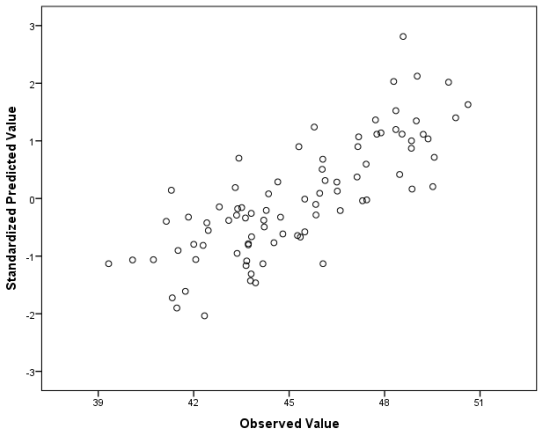
Figure C.1 Examination of Assumptions in TFS 2002: Method One



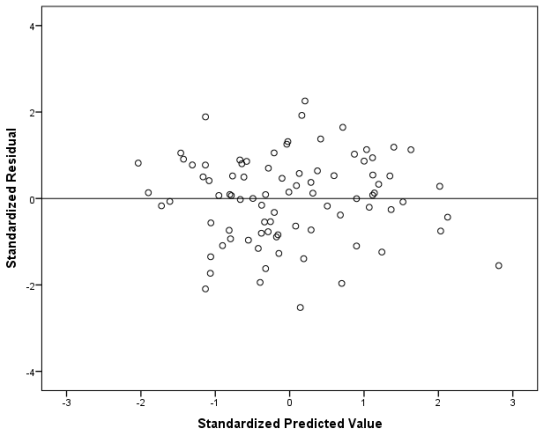
(a) Histogram of Standardized Residuals



(b) Normal P-P Plot of Standardized Residuals

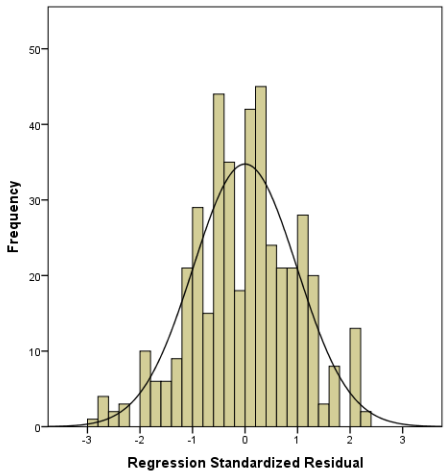


(c) Scatterplot of the Observed and Predicted Civic Engagement Values

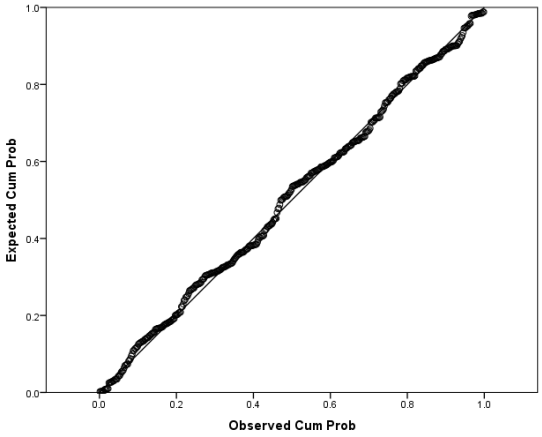


(d) Scatterplot of the Residuals and Predicted Values for Civic Engagement

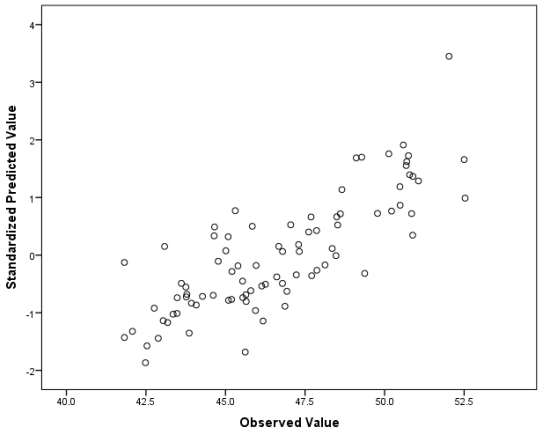
Figure C.2 Examination of Assumptions in TFS 2006: Method One



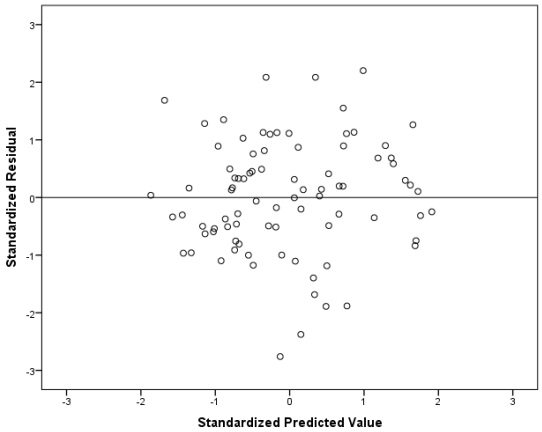
(a) Histogram of Standardized Residuals



(b) Normal P-P Plot of Standardized Residuals

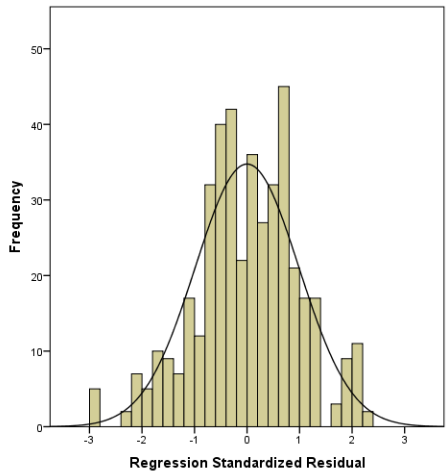


(c) Scatterplot of the Observed and Predicted Civic Engagement Values

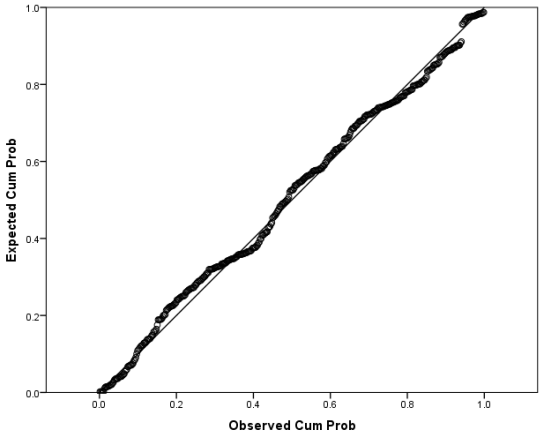


(d) Scatterplot of the Residuals and Predicted Values for Civic Engagement

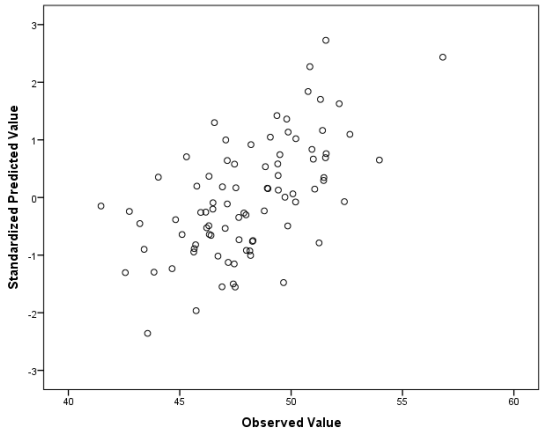
Figure C.3 Examination of Assumptions in CCS 2002: Method One



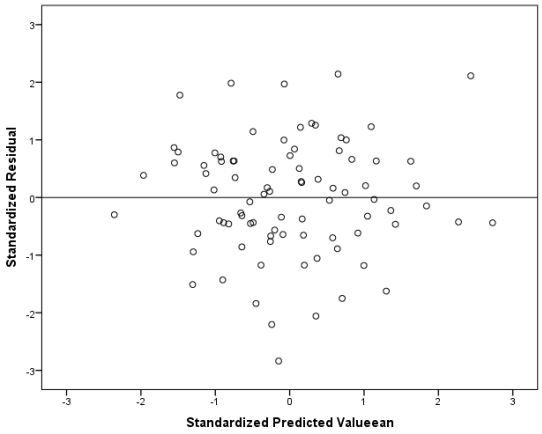
(a) Histogram of Standardized Residuals



(b) Normal P-P Plot of Standardized Residuals

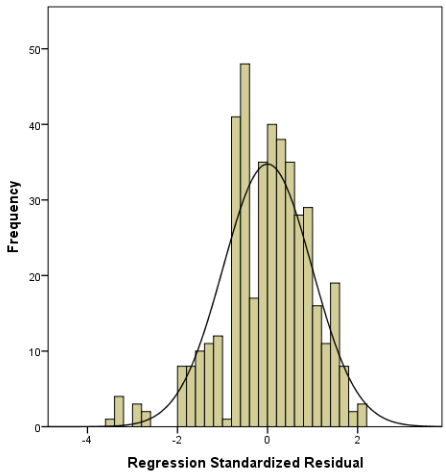


(c) Scatterplot of the Observed and Predicted Civic Engagement Values

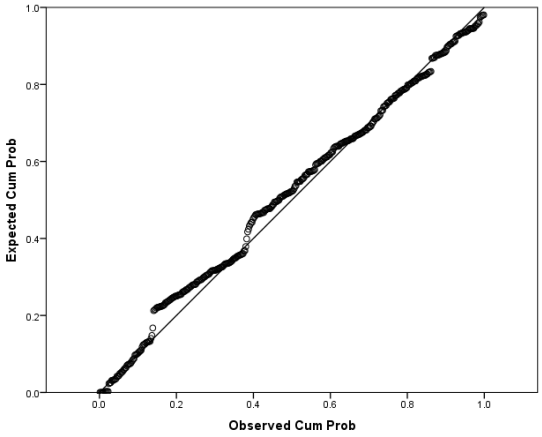


(d) Scatterplot of the Residuals and Predicted Values for Civic Engagement

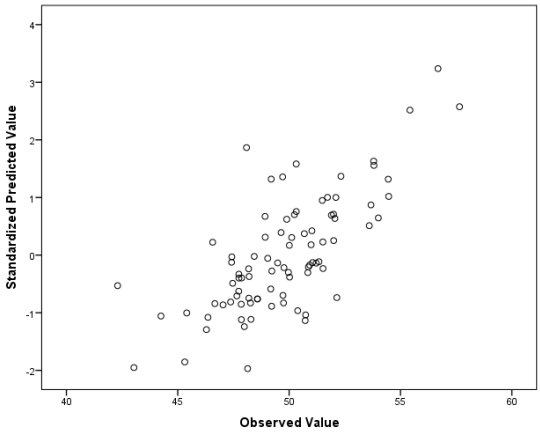
Figure C.4 Examination of Assumptions in CSS 2006: Method One



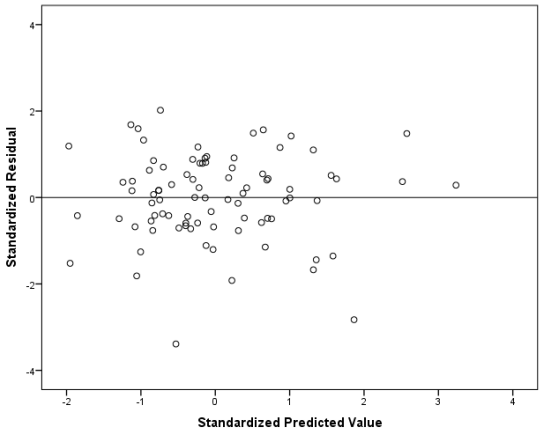
(a) Histogram of Standardized Residuals



(b) Normal P-P Plot of Standardized Residuals



(c) Scatterplot of the Observed and Predicted Civic Engagement Values



(d) Scatterplot of the Residuals and Predicted Values for Civic Engagement

Table C.1 Test of Homogeneity of Level-1 Variance in 2002: Method Two

Chi-square statistic	=	442.71523
Number of degrees of freedom	=	85
P-value	=	0.000

Figure C.5 Q-Q Plot of Mahalanobis Distance and Expected Values in 2002: Method Two

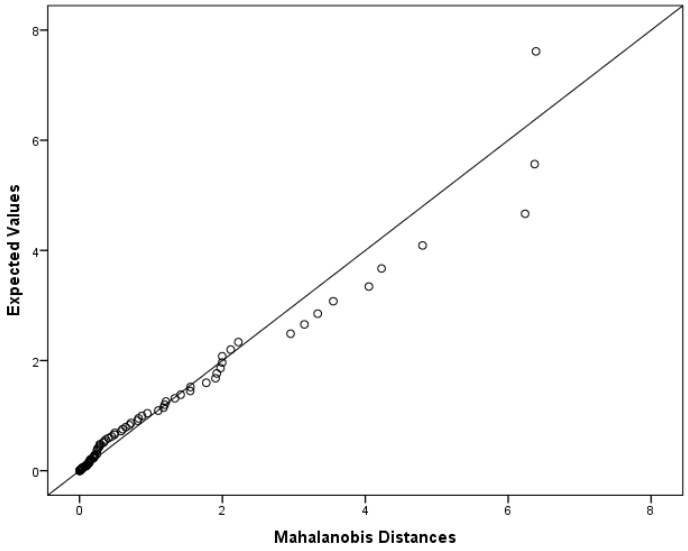


Table C.2 Test of Homogeneity of Level-1 Variance in 2006: Method Two

Chi-square statistic	=	458.21398
Number of degrees of freedom	=	85
P-value	=	0.000

Figure C.6 Q-Q Plot of Mahalanobis Distance and Expected Values in 2006: Method Two

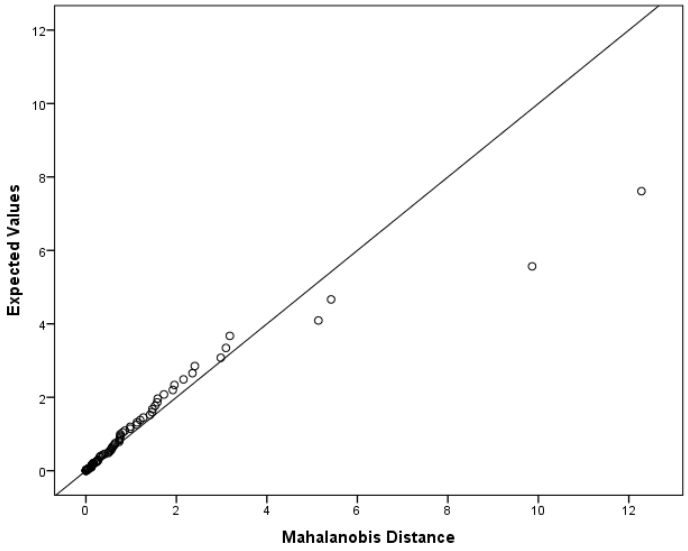


Table C.3 Test of Homogeneity of Level-1 Variance: Method Three

Chi-square statistic	=	359.08615
Number of degrees of freedom	=	89
P-value	=	0.000

Figure C.7 Q-Q Plot of Mahalanobis Distance and Expected Values: Method Three